

TREES, RNA SECONDARY STRUCTURES AND CLUSTER COMBINATORICS

ROBERT J. MARSH AND SIBYLLE SCHROLL

ABSTRACT. We develop and study the structure of combinatorial objects that are a special case of RNA secondary structures. These are generalisations of objects arising from interval exchange transformations generalising those in the Sturmian context. We represent them as labelled edge-coloured trees. We classify and enumerate them and show that a natural subset is in bijection with a set of m -clusters. Furthermore, we interpret a notion of induction used to model generalised interval exchange transformations as a composition of cluster mutations.

1. INTRODUCTION

One of the earliest occurrences of Sturmian words is in a number-theoretic context, in a paper from 1882 by Markov [24] in which he answers a question of Jean Bernoulli. Half a century later, in the early stages of symbolic dynamics, Morse and Hedlund introduced the term *Sturmian*. Sturmian words form a typical example of a minimal set of a dynamical system; this led to a systematic study of their many combinatorial properties. A natural question is how to generalise these special languages to wider classes with less strict combinatorial properties. Sturmian words correspond to interval exchange transformations of two intervals, and one generalisation of Sturmian languages arises from interval exchange transformations of several intervals; see [9]. The combinatorial properties of the associated languages, which were studied in [2], motivated this article.

Another natural generalisation is to consider higher dimensional analogues of interval exchange transformations. The languages arising from this will have a more complex combinatorial structure and it is not immediately obvious what they should look like. In this paper we have taken a complementary approach, generalising the underlying combinatorics and studying it in detail. This generates interesting combinatorics in its own right. It also opens up a possible route to introducing and developing combinatorial languages for higher dimensional exchange transformations. In a future paper [23], currently in preparation, we are taking this approach.

On the other hand, we show that the combinatorics we develop here has a connection with cluster algebras. Cluster algebras were introduced by Fomin and Zelevinsky in [12] in order to model the properties of the dual canonical basis (or global crystal basis) of a quantized enveloping algebra and to study total positivity. As well as bringing greater understanding to these fields, cluster algebras

Date: 11 November 2011.

Key words and phrases. tree, edge-coloured, triangulation, cluster algebra, m -cluster, labelled triangulation, Fuss-Catalan number, enumeration, tree of relations, RNA secondary structure, pseudoknot-free, dissection, polygon, m -angulations, m -divisible, induction, Sturmian word.

2010 Mathematics Subject Classification: Primary: 05C05, 05A15; Secondary: 13F60, 37B10.

This work was supported by the Engineering and Physical Sciences Research Council [grant number EP/G007497/1] and by the Leverhulme Trust in the form of an Early Career Fellowship for Sibylle Schroll.

have been applied with much success in the fields of combinatorics, representation theory and related category theory, as well as other fields. See, for example, the surveys [11, 22].

In particular, cluster algebras led to the introduction [13, 14] of a natural family of simplicial complexes, known as generalized associahedra or cluster complexes, classified by the Dynkin diagrams. The type A case corresponds to the Stasheff polytope. Corresponding polytopes for every Dynkin diagram were constructed in [3]. This family of complexes was generalised further in [10] to give the generalised cluster complexes. By giving natural bijections, we show how the combinatorial objects discussed here relate closely to the generalised cluster complexes of type A .

We now give a more detailed description of the contents of the paper. Given natural numbers k, m we consider the set of labelled m -edge-coloured trees with k vertices, i.e. trees with k vertices, labelled with the numbers $1, 2, \dots, k$, with each edge coloured with one of m possible symbols S_1, S_2, \dots, S_m in such a way that the edges incident with a given vertex all have different colours (see [32, 5.28, p81]); see Figure 4(c) for an example. Such trees have a natural cyclic ordering on their vertices induced by the edge-labelling which we refer to as the circular order.

We show (Section 7) that a form of generalised induction on these trees exists (taking such a tree to another such tree) and that the equivalence classes of its transitive closure are characterised by the circular order. Thus the set of equivalence classes is in bijection with the set of k -cycles in the symmetric group of degree k .

The case $m = 3$ was studied in [2], where the trees we study here were referred to as trees of relations. Furthermore, it was shown in [9] that the induction in the case $m = 3$ induces languages naturally generalising Sturmian languages, arising from interval exchange transformations associated to permutations.

We also consider RNA secondary structures (in the sense of Waterman [35]; see also [28]) on the sequence $(S_1 S_2 \dots S_m)^k$ with the rule that pairings can only be made when the bases at the end-points are equal. Such a secondary structure consists of k vertices, $1, 2, \dots, k$, arranged clockwise around a circle, with m bases given by the m symbols S_1, S_2, \dots, S_m arranged clockwise at each vertex. There is a collection of arcs linking equal bases (with no symbol incident to more than one arc). We call such an RNA secondary structure an *RNA m -diagram of degree k* .

An RNA m -diagram is said to be *noncrossing* if it can be drawn in such a way that no two arcs cross. We say that it is *connected* if every vertex is connected to every other vertex via a sequence of arcs, regarding all of the bases at any given vertex as being connected together. For an example, see Figure 4(a). Diagrams of this form, for the case $m = 3$, were considered in [2]. In Section 2 we give a detailed description of the combinatorial objects we are considering in this article.

We first show that the set of labelled m -edge-coloured trees with k vertices and circular order $(1 \ 2 \ \dots \ k)$ (an induction equivalence class) is in bijection with the set of connected noncrossing RNA m -diagrams of degree k . We then compute the cardinality of each of these sets in three different ways. The first method (Section 3) involves a bijection with the set of rooted diagonal-coloured m -angulations of a regular $(m - 2)k + 2$ -sided polygon, up to rotation. The second method (Section 4) involves a recursion formula along the lines of [35, Thm. 2.1]. The third method (Section 5) involves showing that the sequence whose k th term is the number of connected noncrossing RNA m -diagrams with k vertices is the m -fold convolution of the Fuss-Catalan sequence of degree $m - 1$. This can be regarded as giving an alternative proof of a formula for this term (which is given in [20, §3]).

Since the number of induction equivalence classes of labelled m -edge coloured trees with k vertices is equal to the number of k -cycles in the symmetric group of degree k , and each equivalence class has the same cardinality, we obtain a new way

(Section 6) to count the labelled m -edge-coloured trees on k vertices (this number is known: see [32, 5.28, p124]).

As mentioned above, in Section 7 we study the generalised induction on labelled m -edge coloured trees with k vertices. Finally, in Section 8, we give a description of the generalised induction defined here in terms of $(m - 2)$ -cluster mutation.

2. SOME COMBINATORIAL OBJECTS

We firstly introduce the main combinatorial objects we will be considering in this article: edge-coloured trees, RNA m -diagrams, and m -angulations.

2.1. Trees. As mentioned above, given positive integers k, m , we will consider labelled m -edge-coloured trees on k vertices, i.e. trees whose vertices are labelled $1, 2, \dots, k$ and whose edges are coloured with one of the m symbols S_1, S_2, \dots, S_m in such a way that no two edges incident with the same vertex have the same colour. We will also consider m -edge coloured trees on k vertices (i.e. without any labelling on the vertices). Note that either kind of tree may be *rooted*, i.e. come together with a distinguished vertex. As is usual, if the connectedness assumption is not satisfied, we refer to the corresponding objects as forests. See Figure 4(b) for an example of a (rooted) m -edge-coloured tree and (c) for an example of a labelled m -edge-coloured tree.

Given an m -edge coloured tree \mathcal{S} with k vertices, each symbol S_r determines a map (with the same name) from the set of vertices of \mathcal{S} to itself. A vertex is fixed by S_r unless it is incident with an edge coloured S_r , in which case it is sent to the vertex at the other end of that edge. Let $\sigma_{\mathcal{S}}$ be the composition $S_m S_{m-1} \cdots S_1$: this is a permutation of the vertices of \mathcal{S} . We use the same definition for a labelled m -edge coloured tree, G , obtaining a permutation σ_G in the symmetric group of degree k .

We refer to $\sigma_{\mathcal{S}}$ (respectively, σ_G) as the *circular order* of \mathcal{S} (respectively, G). These maps (for 3 symbols) were considered in [2]. We shall see later (see Lemma 6.1) that the circular order is always a k -cycle.

2.2. RNA secondary structure. The primary structure of an RNA molecule is a linear sequence of nucleotides, each containing a base of one of four types, labelled **A** (adenine), **C** (cytosine), **G** (guanine), or **U** (uracil). The bases may form pairs and such pairing is known as the *secondary structure* of the RNA (see e.g. [31]).

This secondary structure is represented by a *diagram* defined by Waterman in [35] in a key paper initiating the combinatorial study of RNA secondary structure. See also [28, Chapter 2]. The diagram is a graph with vertices $1, 2, \dots, k$ (corresponding to the bases) such that the valency of each vertex is at most 1, represented in the plane by drawing the vertices $1, 2, \dots, k$ in a horizontal line. A pairing between the bases at vertices i and j , with $i < j$, is represented by an arc (i, j) ; its length is said to be $j - i$. The number k is called the *length* of the diagram. For an example, see Figure 1.

A *crossing* is a pair $(i_1, j_1), (i_2, j_2)$ of distinct arcs satisfying $i_1 < i_2 < j_1 < j_2$, and a diagram is said to be *noncrossing* (or *pseudoknot-free*) if it contains no crossings.

Diagrams are often drawn on a circle, using the *Nussinov circle representation* [25]; see Figure 1.

The combinatorics of the secondary structure of RNA was initially developed by Waterman and his collaborators; see [35, 36] and e.g. [37] for a survey of some later work. Others have worked on the subject, e.g. [18]. More recently the combinatorics has also been studied extensively by Reidys; see [28] and references therein.

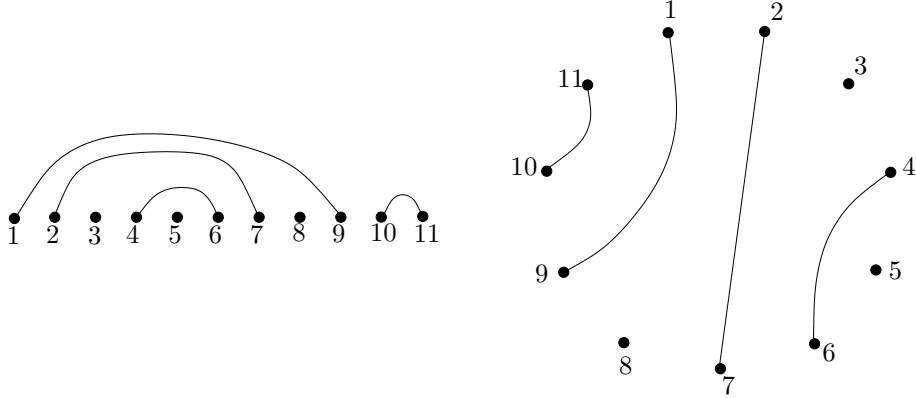


FIGURE 1. A diagram (RNA secondary structure) of length 11 and its Nussinov circle representation.

We remark that RNA secondary structure is also interesting from a topological perspective; see e.g. [26]. Also, the article [27] counts modular diagrams (i.e. having no isolated arcs and with each arc of length at least 4) satisfying a certain noncrossing property, important for a particular model of RNA. RNA secondary structures can be represented using permutations; see [38].

Note that RNA secondary structures can also be considered in a more abstract setting, i.e. with an arbitrary finite set of possible bases and arbitrary pairing rules; see [35, §3] (or [18, §5]). Here we consider such a setup in which there are m bases, S_1, S_2, \dots, S_m . We take the sequence $(S_1 S_2 \cdots S_m)^k$, and the rule that pairings can only be made provided the bases are equal. We define an *RNA m -diagram of degree k* to be a diagram in the above sense, drawn in the Nussinov circle representation, but keeping the bases as labels on the vertices.

Thus there are k vertices $1, 2, \dots, k$, numbered clockwise in a circle. Each vertex contains the bases S_1, S_2, \dots, S_m written in order clockwise, and a collection of arcs connect equal bases at different vertices. A base at a vertex is incident to at most one arc. Such a diagram is *noncrossing* if it can be drawn in such a way that there are no crossings between the arcs. We say that it is *connected* provided there is a path between any two vertices (where moving between bases at a vertex is allowed). See Figure 4(a) for an example of a connected noncrossing RNA 4-diagram of degree 10.

We note that the combinatorics of repeated RNA sequences is also considered in [19, §3] but that a symmetry condition is imposed. We also remark that a bijection between a set of RNA secondary structures and a set of trees of a certain kind is given in [29], but this is not the same as the bijections considered here.

We also remark that the article [1] defines Fuss-Catalan algebras, whose basis diagrams are the same as RNA m -diagrams except that every base is incident with exactly one arc, there is no connectedness condition and the ordering of the bases is reversed in neighbouring vertices except possibly the first and last vertex. In this case, as the name suggests, the number of basis diagrams coincides with a Fuss-Catalan number.

If the bases at each vertex are all amalgamated into a single vertex, a noncrossing RNA m -diagram of degree k gives rise to a generalised version of a tangled diagram in the sense of [4, 5], i.e. in which every vertex has valency at most m (instead of the usual restriction of 2). But note that not every tangled diagram arises in this way; in particular, the diagrams obtained in this way will have at most one arc between any given pair of vertices, no crossings, and no loops or cycles (see Theorem 3.1).

The case we consider has its own combinatorics, which we will investigate here.

2.3. m -angulations. Now let P be an $dk + 2$ -sided regular polygon. A d -diagonal in P is a diagonal joining two vertices of P which divides P into an $dj + 2$ -sided polygon and an $d(n - j) + 2$ -sided polygon for some j where $1 \leq j \leq \lceil \frac{n-1}{2} \rceil$. A maximal collection of d -diagonals of P divides P into $d + 2$ -sided polygons. Such dissections of P are referred to as d -divisible dissections in [34]. Taking $d = m - 2$, such dissections divide P into m -sided polygons, and we shall refer to them here as m -angulations of P .

We say that an m -angulation of P is *diagonal-coloured* if each diagonal in P is coloured with a symbol from the set $\{S_1, S_2, \dots, S_m\}$ in such a way that the colours on the sides of each m -sided polygon in the m -angulation are S_1, S_2, \dots, S_m in clockwise order. We say that it is *rooted* if there is a distinguished m -sided polygon in the m -angulation. We say that it is *m -gon-labelled* if the m -gons are labelled $1, 2, \dots, k$. See Figure 6 for an example of a diagonal-coloured 4-angulation.

3. BIJECTIONS

Our main aim in this section is to compute the number of degree k connected noncrossing RNA m -diagrams. We shall also give a bijective proof that this is the same as the number of labelled m -edge-coloured trees with k vertices and circular order $(k \ k - 1 \ \dots \ 1)$. This is a generalisation of the case $m = 3$ appearing in [2, Sect. 8].

Theorem 3.1. *There is a bijection between the following sets:*

- (I) *The set of degree k noncrossing RNA m -diagrams.*
- (II) *The set of labelled m -edge-coloured forests G with k vertices such that, on writing $G = \sqcup_i G_i$ as a union of connected components, we have the following:*
 - (a) *If $i \neq j$ and $a_1, a_2 \in G_i$, $b_1, b_2 \in G_j$, we cannot have that $a_1 > b_1 > a_2 > b_2$.*
 - (b) *If $a \in G_i$ for some i , then $\sigma_G(a)$ is the maximal vertex of G less than a lying in G_i (or, if no such vertex exists, it is the largest vertex of G lying in G_i).*

Proof. Let Σ be a noncrossing RNA m -diagram of degree k on m symbols as in (I).

Let G be the labelled m -edge-coloured graph with k vertices whose edges given by the arcs of Σ . That is, there is an edge between vertices i and j of G coloured with S_k if and only if there is an arc in Σ between the instances of the symbol S_k in vertices i and j in Σ .

Claim: G is a forest.

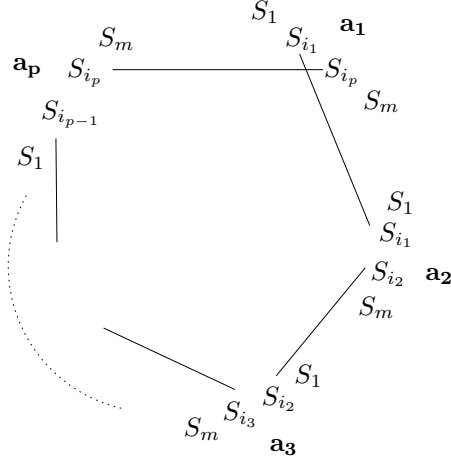
We prove the claim. Suppose, for a contradiction, there is a cycle

$$a_1 \xrightarrow{S_{i_1}} a_2 \xrightarrow{S_{i_2}} \dots \xrightarrow{S_{i_{p-1}}} a_p \xrightarrow{S_{i_p}} a_1$$

in G , and thus a corresponding cycle in Σ . Without loss of generality, we may assume that $i_1 < i_2$. For vertices a, b , we denote by (a, b) the set of vertices c of Σ lying strictly clockwise of a and strictly anticlockwise of b .

Then $a_3 \in (a_2, a_1)$, since the arc in Σ corresponding to the edge in G between a_2 and a_3 cannot cross the arc in Σ corresponding to the edge in G between a_1 and a_2 .

By assumption on Σ , $i_2 \neq i_3$. If $i_2 > i_3$, there can be no path in Σ from the symbol S_{i_3} in vertex a_3 of Σ back to vertex a_1 of Σ without crossings, a contradiction, hence $i_2 < i_3$.

FIGURE 2. A cycle in G leads to a crossing

Repeating this argument, we see that, moving clockwise on Σ from a_1 we meet vertices a_2, a_3, \dots, a_p in order before returning to a_1 , and that $i_1 < i_2 < \dots < i_p$. But then the arc between a_p and a_1 (on symbol S_{i_p} crosses the arc between a_1 and a_2 (on symbol S_{i_1}), since $i_1 < i_p$; see Figure 2. Hence G has no cycles, and must be a tree. The claim is shown.

We next prove that (a) holds. Suppose that $i \neq j$, $a_1, a_2 \in G_i$, $b_1, b_2 \in G_j$, and $a_1 > b_1 > a_2 > b_2$. Then a_1, b_1, a_2, b_2 follow each other anticlockwise around the circle. Then, since a_1, a_2 are in the same connected component of G , there is a path in Σ from a_1 to a_2 , and similarly from b_1 to b_2 . The arrangement of a_1, a_2, b_1 and b_2 implies that these two paths cross, a contradiction. Hence no such arrangement can occur, and (a) is shown.

We next prove that (b) holds. It is enough to prove the following claim:

Claim: Let $a \in G_i$. Then $\sigma_G(a)$ is the next vertex of G_i (considered as a vertex of Σ) anticlockwise on the circle from a .

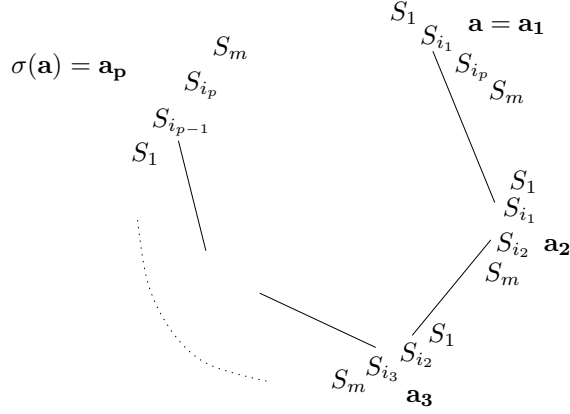
To prove the claim, we note that, by the definition of σ_G , $\sigma_G(a)$ is connected to a by a path in G :

$$a = a_1 \xrightarrow{S_{i_1}} a_2 \xrightarrow{S_{i_2}} \dots \dots \xrightarrow{S_{i_{p-1}}} a_p = \sigma_G(a),$$

where $i_1 < i_2 < \dots < i_{p-1}$. Furthermore, a_1 is not incident with any arc with symbol S_r for $r < i_1$, a_p is not incident with any arc with symbol S_r for $r > i_{p-1}$, and, for $2 \leq j \leq p-1$, a_j is not incident with any symbol S_r for $i_{j-1} < r < i_j$.

The existence of the above path implies that $\sigma_G(a)$ lies in G_i . Since $i_1 < i_2 < \dots < i_p$ and there are no crossings, the path must go clockwise around the circle; see Figure 3. The conditions above and the fact there are no crossings imply that no vertex in $(\sigma_G(a), a)$ has an arc with a vertex in $(a, \sigma_G(a))$ or with a or $\sigma_G(a)$, so these vertices are connected only amongst themselves. It follows that they do not lie in G_i and we see that (II)(b) holds. Thus G is a labelled m -edge-coloured forest satisfying (II).

Conversely, suppose that we have a labelled m -edge-coloured forest with k vertices satisfying (II). Let Σ be the RNA m -diagram of degree k with an arc joining S_r in vertex i with S_r in vertex j if and only if there is an edge in G between vertices i and j coloured with symbol S_r . We must check that Σ can be drawn with no crossing arcs, i.e. that it is noncrossing.

FIGURE 3. The part of Σ between a and $\sigma_G(a)$.

We do this by induction on the number of vertices. Suppose first that G has more than one connected component, i.e. that G is not connected. By induction, each component G_i corresponds to a noncrossing RNA m -diagram (on the vertices of G_i).

Suppose that we had $a_1 > a_2 \in G_i$ and $b_1 > b_2 \in G_j$ for two distinct components G_i and G_j , with arcs between a_1 and a_2 and b_1 and b_2 which cross in Σ . Then, going around the circle anticlockwise, starting at vertex k , we must encounter a_1, b_1, a_2, b_2 in order, or b_1, a_1, b_2, a_2 in order. Swapping G_i and G_j if necessary, we can assume we are in the first case. But then $a_1 > b_1 > a_2 > b_2$, contradicting (II)(a). Hence Σ is noncrossing.

So we are reduced to the case in which G has exactly one connected component, i.e. G is connected. Suppose that vertex k is incident with edges e_1, e_2, \dots, e_d in G , coloured with symbols $S_{r_1}, S_{r_2}, \dots, S_{r_d}$ where $r_1 < r_2 < \dots < r_d$. Let the endpoints of these edges (other than k) be v_1, v_2, \dots, v_d . Removing vertex k from G leaves precisely d trees T_1, T_2, \dots, T_d containing vertices v_1, v_2, \dots, v_d respectively. By (b), we know that $\sigma_G = S_m S_{m-1} \dots S_1$ induces the permutation $(k \ k-1 \ \dots \ 1)$ on the vertices of G .

We apply σ_G to vertex k , and then repeatedly apply σ_G . By its definition, each application of σ_G corresponds to following a certain path through G , i.e. passing along the edges corresponding to the symbols in the sequence S_1, S_2, \dots, S_m in that order, when such incident edges exist. Since the edge e_1 has symbol S_{r_1} , and no edge incident with k has smaller symbol, it follows that, after the first application of σ_G , we obtain vertex $k_1 := k-1$ in T_1 .

Since σ_G is a k -cycle, after repeated application of σ_G , we must leave T_1 . Suppose that k_2 is the number of the first vertex reached outside T_1 . Since r_2 is the minimum number of a symbol adjacent to k greater than r_1 , k_2 will lie in T_2 . Repeating this argument, we will obtain $k > k_1 > k_2 > \dots > k_d \geq 1$ such that vertices $k_{i+1} + 1, \dots, k_i$ lie in tree T_i for $i = 1, 2, \dots, d-1$. At the final step, the first vertex reached on leaving T_d must be k . Since σ_G is a k -cycle, all vertices must have been visited.

Let $k_{d+1} = 0$. It follows from the above that tree T_i contains precisely vertices $k_{i+1} + 1, \dots, k_i$ for each i . Thus, the numbering of the vertices of G is first the vertices of T_d in some order, then the vertices of T_{d-1} in some order, then the vertices of T_{d-2} , and so on, ending with the vertices of T_1 and then finally k .

Each T_i will correspond (by the inductive hypothesis) to a noncrossing RNA m -diagram on its vertices. Thus the vertices v_1, v_2, \dots, v_d in G will be numbered in decreasing order. The arcs in Σ from k to these vertices are numbered by symbols $S_{r_1}, S_{r_2}, \dots, S_{r_d}$, respectively, with $r_1 < r_2 < \dots < r_d$. It follows these arcs do not cross each other or any of the other arcs in Σ . See Figure 4(a) for an example, where $v_1 = 9$, $r_1 = 1$, and $v_2 = 8$, $r_2 = 4$. Hence, Σ is noncrossing and thus an object in (I) as required.

It is clear that the two maps we have constructed are inverse to each other, so the theorem is proved. \square

The following lemma follows easily from the definitions.

Lemma 3.2. *A noncrossing RNA m -diagram is connected if and only if the corresponding labelled m -edge-coloured forest is connected, i.e. is a tree.*

Remark 3.3. (1) Since a forest on k vertices is connected if and only if it has exactly $k - 1$ edges, a noncrossing RNA m -diagram with k vertices is connected if and only if it has $k - 1$ arcs.

(2) In the connected case, the circular order of G is just the permutation $(k \ k - 1 \ \dots \ 1)$ and we have a bijection between the following sets:

The set of connected noncrossing RNA m -diagrams of degree k .

\updownarrow

The set of labelled m -edge-coloured trees with k vertices and circular order $(k \ k - 1 \ \dots \ 1)$.

The labelling on the vertices for a tree in the latter set is determined by a distinguished vertex, that labelled k , say, since σ_G then determines the labels on all the other vertices. We thus have:

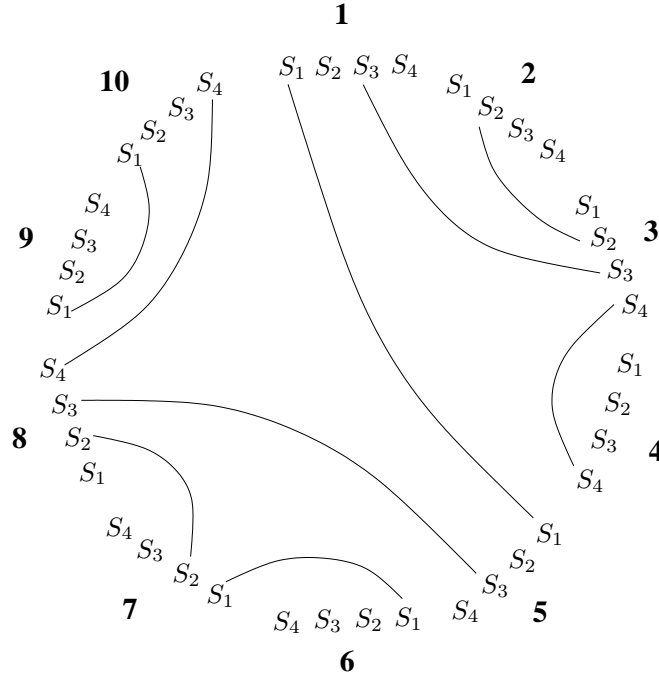
Corollary 3.4. *The bijection in Theorem 3.1 induces a bijection between the following sets:*

- (a) *The set of connected noncrossing RNA m -diagrams of degree k .*
- (b) *The set of rooted m -edge-coloured trees with k vertices.*
- (c) *The set of labelled m -edge-coloured trees with k vertices and circular order $(k \ k - 1 \ \dots \ 1)$.*

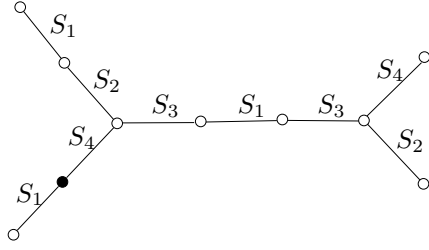
See Figure 4(a)-(c) for an illustration of this bijection, i.e. a particular connected noncrossing RNA m -diagram of degree 10 on 4 symbols and the corresponding objects from (b) and (c) above.

Remark 3.5. Following [39, Sect. 5], define a noncrossing RNA m -diagram of degree k to be *saturated* if it is maximal in the sense that adding any further arcs would destroy the noncrossing property. Note that saturated RNA m -diagrams have also been studied in [7]. By Remark 3.3(1), a noncrossing RNA m -diagram is connected if and only if it has $k - 1$ arcs. Thus, if any more arcs were added, it would no longer be noncrossing. It follows that any noncrossing RNA m -diagram which is connected is necessarily saturated. Finally, we note that the noncrossing RNA 3-diagram of degree 5 in Figure 5 is saturated but not connected.

In what follows, we need a bijection between the set of diagonal-coloured m -angulations of an $(m - 2)k + 2$ -sided regular polygon up to rotation and the set of m -edge-coloured trees with k vertices. Such a bijection in the unlabelled case (between m -angulations up to rotation and plane $(m - 1)$ -ary trees) is well-known; see, for example, [32, Sect. 6.2]. Here we need a labelled version, and for convenience we give some details of how this can be done. We also note that the case $m = 3$ appears in [23].



(a) A connected noncrossing RNA 4-diagram of degree 10.



(b) A rooted 4-edge-coloured tree with 10 vertices.

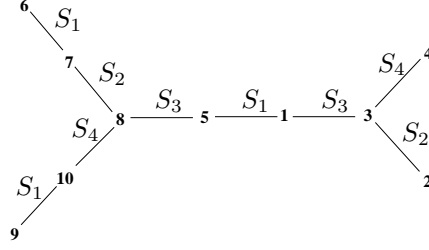
(c) A labelled 4-edge-coloured tree with 10 vertices and circular order $(10\ 9 \dots 1)$.

FIGURE 4. Objects corresponding to each other under the bijections in Corollary 3.4.

Theorem 3.6. *There is a bijection between the following sets:*

- (a) *The set of m -edge-coloured trees with k vertices.*
- (b) *The set of diagonal-coloured m -angulations of an $(m-2)k+2$ -sided regular polygon up to rotation.*

Proof. Given a diagonal-coloured m -angulation of an $(m-2)k+2$ -sided polygon, take the dual graph, with a vertex in the middle of each m -sided polygon in the m -angulation and an edge between two vertices coloured S_i whenever the corresponding polygons share an edge in the dissection coloured S_i . Since the edges of a polygon in the m -angulation all have distinct colours, the same is true for the edges in the dual graph incident with a given vertex. Since there are no internal vertices in the m -angulation, there can be no cycles in the dual graph. It is clearly connected, hence it is a tree. By construction, it is m -edge-coloured and has k vertices.

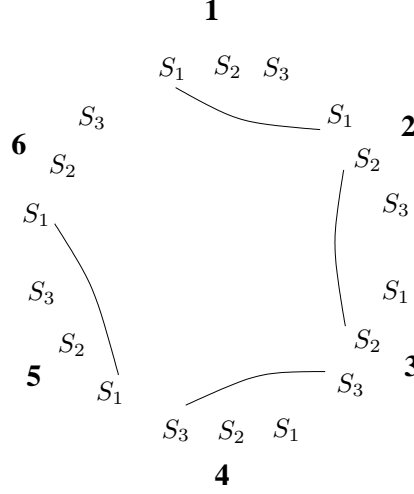


FIGURE 5. A noncrossing RNA 3-diagram of degree 5 which is saturated but not connected.

Conversely, given an m -edge-coloured tree with k vertices, first complete it as follows. For each vertex, add extra edges coloured with all the symbols not already colouring the edges incident with the vertex. The other end of each additional edge added is a new vertex of valency 1 and is called a boundary vertex. Each original vertex has valency m in the completed tree and is referred to as an interior vertex.

Next, map the completed tree to the plane so that the edges around each vertex are coloured with the symbols S_1, S_2, \dots, S_m in clockwise order, with all edges of unit length, and so that the edges around any given vertex are equally spaced, with an angle of $2\pi/m$ between successive edges. Note that vertices and edges may overlap. It is clear that any two such maps will be the same up to a rotation.

Then, each interior vertex determines a regular m -gon in the plane, with the mid-points of the edges in the polygon given by the mid-points of the m edges adjacent to the given vertex in the original tree. The edges in the polygon are given the same colours as the edges in the completed tree that they cross. The collection of regular m -gons obtained (which may also overlap) forms an m -cluster in the sense of [16] (note that this is not the more modern notion of cluster introduced by Fomin-Zelevinsky [12]). Then, as in [16, Sect. 7] there is an (orientation-preserving) homeomorphism of the plane taking the m -cluster to a (diagonal-coloured) m -angulation of P . The dissection obtained is well-defined up to a rotational symmetry of P (by the above remark concerning the map of the completed tree into the plane).

It is clear that this map is the inverse of the map above so the result is proved. \square

We remark that the completion of trees considered here is essentially (a generalisation of) the standard completion of binary trees (i.e. a bijection from (c) to (d) in [32, Ex. 6.19]).

Corollary 3.7. *There is a bijection between the following sets:*

- (a) *The set of labelled m -edge-coloured trees of relations with k vertices and circular order $(k \ k - 1 \ \dots \ 1)$.*
- (b) *The set of rooted diagonal-coloured m -angulations of an $(m - 2)k + 2$ -sided regular polygon up to rotation.*

Proof. We use the above bijection. Given a labelled m -edge-coloured tree with k vertices, forget the vertex-labelling to obtain an m -edge-coloured tree with k vertices, and take the corresponding diagonal-coloured m -angulation together with

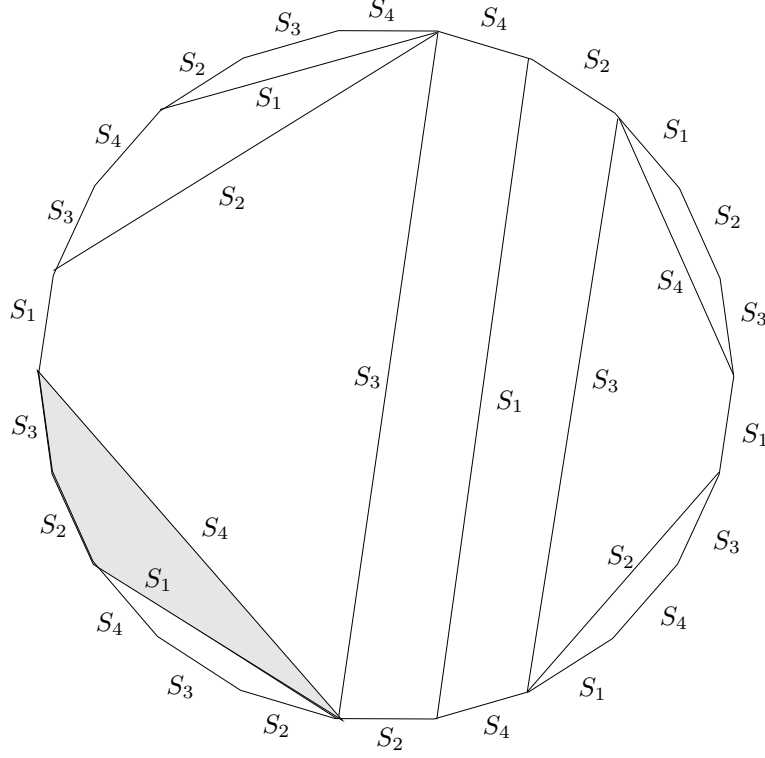


FIGURE 6. 4-angulation of a regular 22-sided polygon as in Corollary 3.7(b) corresponding to the labelled 4-edge-coloured tree with 10 vertices in Figure 4(c) via the bijection in Corollary 3.7. The shaded region is the root m -gon.

the distinguished m -sided polygon corresponding to the vertex in the original tree labelled k .

Conversely, given a rooted diagonal-coloured m -angulation of an $(m-2)k+2$ -sided regular polygon (up to rotation), ignore the root and take the corresponding m -edge-coloured tree with k vertices, \mathcal{S} . The root gives a distinguished vertex of the tree, which we label k . Then label the vertex $\sigma_{\mathcal{S}}^i(k)$ with $k-i$, for $i = 1, 2, \dots, k$. It is clear that this numbering gives a labelling of the m -edge-coloured tree giving it circular order $(k \ k-1 \ \dots \ 1)$. \square

See Figure 6 for the rooted diagonal-coloured 4-angulation of an $(4-2) \cdot 10 + 2 = 22$ -sided regular polygon (up to rotation) corresponding to the labelled 4-edge-coloured tree with 10 vertices and circular order $(10 \ 9 \ \dots \ 1)$ in Figure 4(c).

Corollary 3.8. *The cardinality of each of the following sets:*

- (a) *The set of labelled m -edge-coloured trees with k vertices and circular order $(k \ k-1 \ \dots \ 1)$;*
- (b) *The set of connected noncrossing RNA m -diagrams of degree k ;*

is given by

$$T_{k,m} = \frac{m}{(m-2)k+2} \binom{(m-1)k}{k-1}.$$

Proof. Both sets have the same cardinality by Corollary 3.4 and, by Corollary 3.7, they have the same cardinality as the set of rooted diagonal-coloured m -angulations of an $(m-2)k+2$ -sided regular polygon up to rotation. The number $S_{k,m}$ of such

m -angulations without a root, with no labelling of diagonals and ignoring rotational equivalence is well-known (see e.g. [17]). Let C_k^m be the k th Fuss-Catalan number of degree m :

$$C_k^m = \frac{1}{k} \binom{mk}{k-1} = \frac{1}{(m-1)k+1} \binom{mk}{k}.$$

Then

$$S_{k,m} = C_k^{m-1} = \frac{1}{(m-2)k+1} \binom{(m-1)k}{k}.$$

Since there are k m -sided polygons in an m -angulation, there are k possibilities for the root. There are m possibilities for a labelling (once one diagonal is coloured, all other diagonals in the m -angulation have determined colours using the rule that each m -gon must have its edges coloured S_1, S_2, \dots, S_m clockwise around the boundary. Each orbit of diagonal-coloured rooted dissections under the action of the rotation group of the polygon contains $(m-2)k+2$ elements (the number of sides of P). Hence, we have:

$$T_{k,m} = \frac{kmS_{k,m}}{(m-2)k+2}$$

and the result follows. \square

Example 3.9. For $m = 3, 4, 5, 6$, the first few values of $T_{k,m}$ are given in the following table:

k	0	1	2	3	4	5	6
$T_{k,3}$	1	1	3	9	28	90	297
$T_{k,4}$	1	1	4	18	88	455	2448
$T_{k,5}$	1	1	5	30	200	1425	10626
$T_{k,6}$	1	1	6	45	380	3450	32886

For $m = 3$, we have that:

$$T_{k,3} = \frac{3}{k+2} \binom{2k}{k-1} = C(k+1) - C(k),$$

where $C(r) = \frac{1}{r+1} \binom{2r}{r}$ is the r th Catalan number. This case is covered in [2, Prop. 7.5], and is sequence A071724 in [30]. For $m = 4$, we have:

$$T_{k,4} = \frac{4}{2k+2} \binom{3k}{k-1},$$

which is sequence A006629 in [30]. The cases $m = 5, 6$ do not appear in [30].

Remark 3.10. Let $\pi \in \mathfrak{S}_k$. In 3.1-3.8, we could consider noncrossing RNA m -diagrams on vertices $\pi(1), \pi(2), \dots, \pi(k)$ (clockwise around the circle). Then the results 3.1-3.8 hold, with the circular order $(k \ k-1 \ \dots \ 1)$ replaced with the circular order $\sigma = (\pi(k) \ \pi(k-1) \ \dots \ \pi(1))$ satisfying $\sigma(\pi(i)) = \pi(i-1)$.

In particular, parts (a) and (b) of Theorem 3.1 become:

- (a) If $\pi(a_1), \pi(a_2) \in G_i$ and $\pi(b_1), \pi(b_2) \in G_j$ for $i \neq j$, then we cannot have $a_1 > b_1 > a_2 > b_2$.
- (b) If $\pi(a) \in G_i$ for some i , then $\sigma_G(\pi(a))$ is the vertex $\pi(a')$ with a' less than a maximal such that $\pi(a') \in G_i$ (or, if no such vertex exists, it is $\pi(a')$ where a' is maximal such that $\pi(a')$ lies in G_i).

Furthermore, the number of connected labelled m -edge-coloured trees with k vertices and fixed circular order σ given by a k -cycle is given by the formula for $T_{k,m}$ in Corollary 3.8.

4. LINK WITH FUSS-CATALAN COMBINATORICS

In this section, we show that by restricting degree k noncrossing RNA m -diagrams appropriately, we get a bijection with the set $(m-2)$ -clusters of type A_{k-1} (in the sense of [10]) or the set of m -angulations of an $(m-2)k+2$ -sided regular polygon. Since the number of either of these is known (a Fuss-Catalan number), this, together with an appropriate bijection, gives us an alternative way of counting the total number of RNA m -diagrams, i.e. an alternative proof of Corollary 3.8.

We also consider here m -angulations of a fixed polygon, not up to rotation. Note that such an m -angulation can be considered as an m -angulation of a polygon together with a marked edge. The marked edge (which can be considered the base of the polygon) indicates how the m -angulation can be rotated onto the fixed polygon we are interested in.

We now state the result giving the bijections, emphasizing that the bijection between the sets in parts (3) and (4) is essentially a variation of Theorem 3.6 (which uses [32, Sect. 6.2]), and that the bijection between (4) and (7) is in [10, Sect. 5.1]. The set in (5) is a staging post, and the sets in (6) and (7) are included for interest. Thus the main content is linking the sets in (1) and (2) with each other and with the other sets listed.

Theorem 4.1. *The following sets are all in bijection.*

- (1) *The set of degree $k+1$ connected noncrossing RNA m -diagrams with $(k+1)st$ vertex containing S_1 only;*
- (2) *The set of labelled m -edge-coloured trees G with $k+1$ vertices and circular order $(k+1 \ k \ \cdots \ 1)$, with $k+1$ a vertex of valency 1 and the edge adjacent to it coloured S_1 ;*
- (3) *The set of rooted m -edge-coloured trees with $k+1$ vertices and only one edge, coloured S_1 , adjacent to the root;*
- (4) *The set of m -angulations of a fixed $(m-2)k+2$ -sided regular polygon;*
- (5) *The set of rooted m -edge-coloured trees with k vertices and no edge coloured S_1 adjacent to the root;*
- (6) *The set of rooted complete $(m-1)$ -plane trees with k internal vertices;*
- (7) *The set of $(m-2)$ -clusters (in the sense of [10]) of type A_{k-1} , i.e. maximal simplices of the $(m-2)$ -cluster complex of type A_{k-1} as considered in [10].*

Proof. (1) \leftrightarrow (2): This is a restriction of the bijection in Remark 3.3. The fact that vertex $k+1$ contains only the symbol S_1 corresponds to the fact that vertex $k+1$ in the tree has only one edge incident with it, coloured S_1 .

(2) \leftrightarrow (3): To go from an object in (2) to an object in (3), take vertex $k+1$ as the root and forget the vertex-labelling. For the inverse map, the condition on the circular order, σ_G , in (2) determines the labelling, starting from $k+1$ corresponding to the root.

(3) \leftrightarrow (4): Restricting the bijection in Theorem 3.6 gives a bijection between the set of objects in (3) and the set of rooted diagonal-coloured m -angulations of an $(m-2)(k+1)+2$ -sided regular polygon up to rotation such that the root m -gon has only one edge incident with another m -gon and that edge is coloured S_1 .

Given such a rooted diagonal-coloured m -angulation of an $(m-2)(k+1)+2$ -sided regular polygon \mathcal{M} up to rotation, we remove the root m -gon to give (via an orientation-preserving homeomorphism) a rooted diagonal-coloured m -angulation of an $(m-2)k+2$ -sided regular polygon. We mark the edge (which is coloured S_1) and forget the diagonal labelling to obtain an m -angulation of a fixed $(m-2)k+2$ -sided regular polygon, i.e. an object as in (4). Note that, because of the marked edge, the m -angulation we obtain is independent of the choice of \mathcal{M} up to rotation.

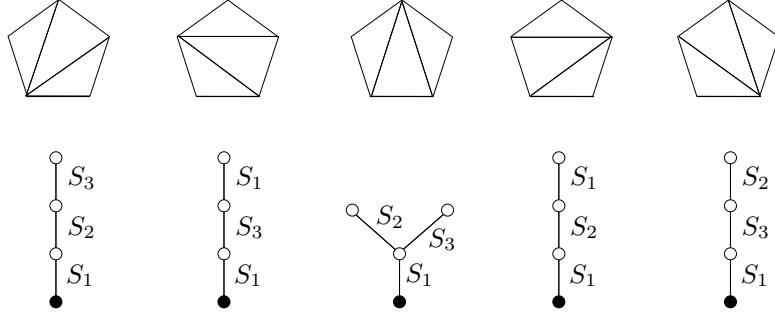


FIGURE 7. The case $k = m = 3$ of the bijection between (3) and (4) in Theorem 4.1.

To go back, given an m -angulation \mathcal{M}' of a fixed $(m-2)k+2$ -sided regular polygon, we add a new m -gon M onto the boundary of \mathcal{M}' with one boundary edge incident with the marked edge of \mathcal{M}' . This, via an orientation-preserving homeomorphism, gives an m -angulation of an $(m-2)(k+1)+2$ -sided regular polygon, which we consider up to rotation. We colour the marked edge of \mathcal{M}' with S_1 . This uniquely determines a diagonal-colouring of the rest of the m -angulation (since each m -gon must have its boundary edges coloured S_1, S_2, \dots, S_m clockwise, by definition). It is clear that this gives an inverse to the above map.

Figure 7 gives an explicit bijection between the set of objects in (3) and the set of objects in (4) in the case $k = m = 3$. We always draw the triangulation of the pentagon with the marked edge at the base.

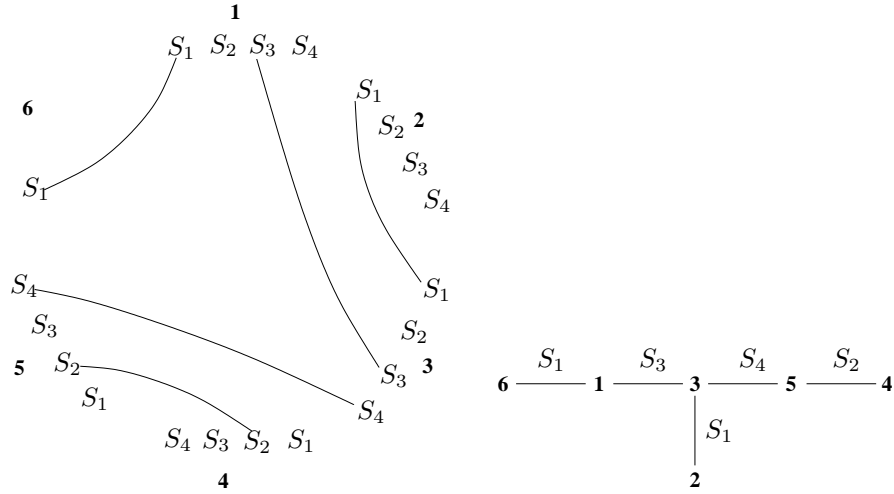
(3) \leftrightarrow (5): Choose an object in (3), and let e be the edge coloured S_1 , incident with the root. Let the new root be the other vertex incident with e , then delete the old root and the edge e . We obtain a new rooted m -edge-coloured tree with k vertices and S_1 not colouring any edge incident with the new root. To go back, given an object in (5), add a new edge incident with the root and colour it S_1 . The other end is a new vertex which becomes the new root.

(5) \leftrightarrow (6): Given an object in (5), first complete it as in the second paragraph of Theorem 3.6. Embed it in plane using the cyclical order $S_1, S_2, \dots, S_m, S_1$ of the edges around each vertex. Keep the root unchanged, and delete the edge incident with the root in the completed tree which is coloured with S_1 . Remove all the colours on the edges. We obtain in this way a rooted completed plane $(m-1)$ -tree.

To go back: given an object in (6), let the *parental* edge of a vertex be the edge between it and its parent. Regard the root as having a (non-existent) parental edge labelled S_1 for these purposes. Then, starting with the root and working inductively, given a vertex with parental edge coloured S_j , colour the edges linking it to its children clockwise from the parental edge in the order: $S_{j+1}, S_{j+2}, \dots, S_m, S_1, \dots, S_{j-1}$. Finally, remove all the leaves.

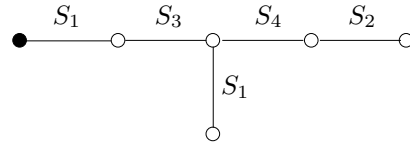
(4) \leftrightarrow (7): See [10, Sect. 5.1]. \square

Figure 8 gives examples of objects corresponding to each other under the above bijections. For (7), we use the numbering of the vertices of the dodecagon anticlockwise, P_1, P_2, \dots, P_{12} , from the root edge at the bottom of part (4) of the figure. Following [10, Sect. 5.1], we see that the 2-snake consists of diagonals linking pairs of vertices (P_1, P_{10}) , (P_{10}, P_3) , (P_3, P_8) and (P_8, P_5) , corresponding to the negative simple roots $-\alpha_1^1$, $-\alpha_2^1$, $-\alpha_3^1$ and $-\alpha_4^1$, respectively. This allows the corresponding 2-cluster to be read off the 4-angulation as in [10, Sect. 5.1]. Here $\alpha_i + \dots + \alpha_j$ is denoted by α_{ij} .

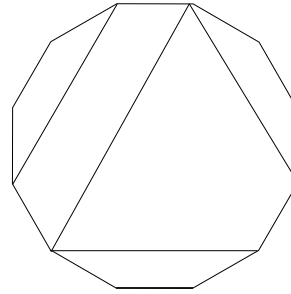


8.1: A degree 6 connected noncrossing RNA m -diagram with 6th vertex containing S_1 only.

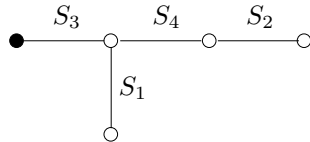
8.2: A labelled 4-edge-coloured tree with 6 vertices and circular order (6 5 4 3 2 1) with 6 a vertex of valency 1 connected to an edge with symbol S_1 .



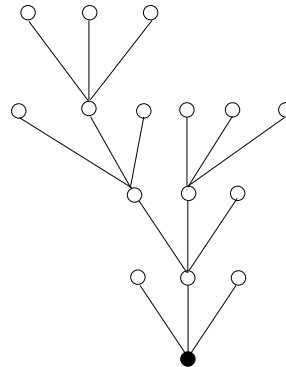
8.3: A rooted 4-edge-coloured tree with 6 vertices and only one edge, coloured S_1 , adjacent to the root.



8.4: An m -angulation of an $(m-2)k+2 = 12$ -sided regular polygon (the marked edge is the one at the bottom of the diagram).



8.5: A rooted 4-edge-coloured tree with 5 vertices and no edge coloured S_1 adjacent to the root.



8.6: A rooted complete ternary plane tree with 5 internal vertices.

$\alpha_1^2, \alpha_{14}^2, \alpha_4^2, \alpha_{34}^1$

8.7: A 2-cluster of type A_4

FIGURE 8. Objects corresponding to each other under the bijections in Theorem 4.1.

As previously remarked, the number of objects in (4) is known (see e.g. [17]) and is given by the Fuss-Catalan number C_k^{m-1} . So we have:

Theorem 4.2. *The number of objects in (1) and (2) is*

$$S_{k,m} = C_k^{m-1} = \frac{1}{k} \binom{(m-1)k}{k-1} = \frac{1}{(m-2)k+1} \binom{(m-1)k}{k}.$$

The number of objects in (2) for the case $m = 3$ is counted in [2, Prop. 7.5].

Example 4.3. For $m = 3, 4, 5, 6$, the values of $S_{k,m}$ for small k are given in the following table for convenience:

k	0	1	2	3	4	5	6
$S_{k,3}$	1	1	2	5	14	42	132
$S_{k,4}$	1	1	3	12	55	273	1428
$S_{k,5}$	1	1	4	22	140	969	7084
$S_{k,6}$	1	1	5	35	285	2530	23751

See [30], sequences A000108, A120588. The case $m = 4$ is sequence A001764, the case $m = 5$ is sequence A002293 and the case $m = 6$ is A002294.

We remark that different sequences generalising the Catalan numbers appear in the enumeration of RNA m -diagrams of various kinds in [33]. See also [8]. Catalan numbers also appear in the enumeration of k -noncrossing RNA secondary structures in [21].

In [35, Thm. 2.1], a recursion formula for the number of RNA secondary structures (i.e. diagrams in the sense of [28, Chap. 2]) is given. The main technique is to partition the diagrams on n points according to whether (a) the point n at one end of the diagram is (i) incident with an arc or (ii) not incident, and then, in case (ii), according to (b) the other end point of the arc. We now use a similar technique to get a recursion-type formula for connected noncrossing RNA m -diagrams of degree k .

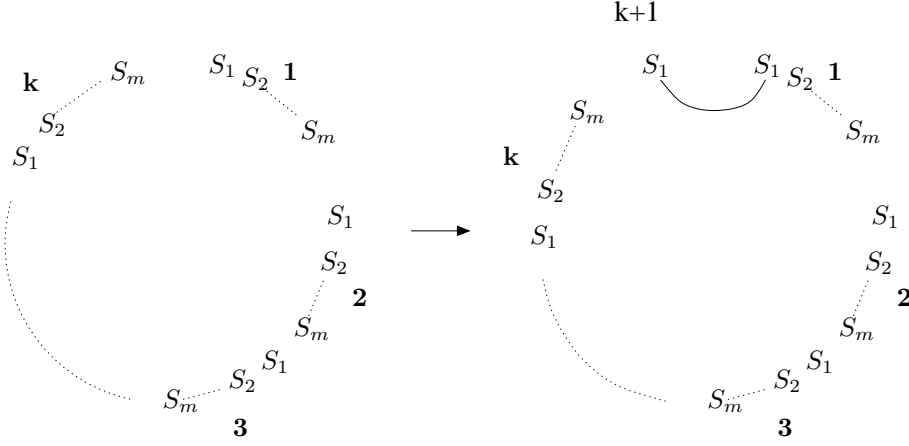
This will give a quadratic formula for $T_{k,m}$ (the number of connected noncrossing RNA m -diagrams of degree k , or the number of labelled m -edge-coloured trees with k vertices and circular order $(k \ k-1 \ \dots \ 1)$) in terms of the $S_{v,m}$ (the number of noncrossing RNA m -diagrams of degree $v+1$ with $(v+1)$ st vertex containing S_1 only, or the number of labelled m -edge-coloured trees with $v+1$ vertices, circular order $(v+1 \ v \ \dots \ 1)$ and $v+1$ a vertex of valency 1 and edge incident to it coloured S_1) with v varying. This allows us to give an alternative proof of the formula for $T_{k,m}$ (Corollary 3.8).

Lemma 4.4. *There is a bijection between the set of connected noncrossing RNA m -diagrams of degree k and the disjoint union of*

- (a) *The set of connected noncrossing RNA m -diagrams of degree $k+1$ with vertex $k+1$ having symbol S_1 only and*
- (b) *The disjoint union (for $v = 2, \dots, k$) of the sets of pairs with the first consisting of a connected noncrossing RNA m -diagram of degree v with vertex v having symbol S_1 only and the second element consisting of a connected noncrossing RNA m -diagram with degree $k-v+2$, vertex $k-v+2$ having symbol S_1 only.*

Proof. Partition the set of connected noncrossing RNA m -diagrams Σ of degree k by

- (a) whether there is an arc in Σ incident with S_1 in vertex 1 and
- (b) if there is such an arc, by the vertex v , $2 \leq v \leq k$, of its other end.

FIGURE 9. The case with no arc incident with S_1 in vertex 1.

The set of structures with no such arc is in bijection with the set of connected noncrossing RNA m -diagrams of degree $k + 1$ with vertex $k + 1$ having symbol S_1 only (the bijection adds a new vertex, $k + 1$ with symbol S_1 only and an arc from S_1 in Σ , vertex 1, to vertex $k + 1$). See Figure 9.

Given a structure Σ with an arc incident with S_1 in vertex 1, we obtain a connected noncrossing RNA m -diagram of degree v with vertex v having S_1 only as a symbol, by restricting to vertices $1, \dots, v - 1$ and symbol S_1 of vertex v .

We obtain a connected noncrossing RNA m -diagram of degree $k - v + 2$ with only symbol S_1 in degree $k - v + 2$ by considering the remainder of Σ not lying in the restriction above, adding the symbol S_1 to the vertex v , renumbering the vertices $1, \dots, k + 1 - v$ and then following the same procedure as in the case when there is no arc incident with S_1 in vertex 1. See Figure 10.

It is clear that this pair determines Σ , and the Lemma follows. \square

Corollary 4.5. *The number of connected noncrossing RNA m -diagrams of degree k on m symbols is given by*

$$T_{k,m} = \sum_{v=1}^k C_{v-1}^{m-1} C_{k+1-v}^{m-1} = \sum_{v=1}^k S_{v-1,m} S_{k+1-v,m}.$$

Proof. This follows immediately from Lemma 4.4, with the first term ($v = 1$) corresponding to the case where there is no arc incident with symbol S_1 in vertex 1 of the connected noncrossing RNA m -diagram structure of degree k . \square

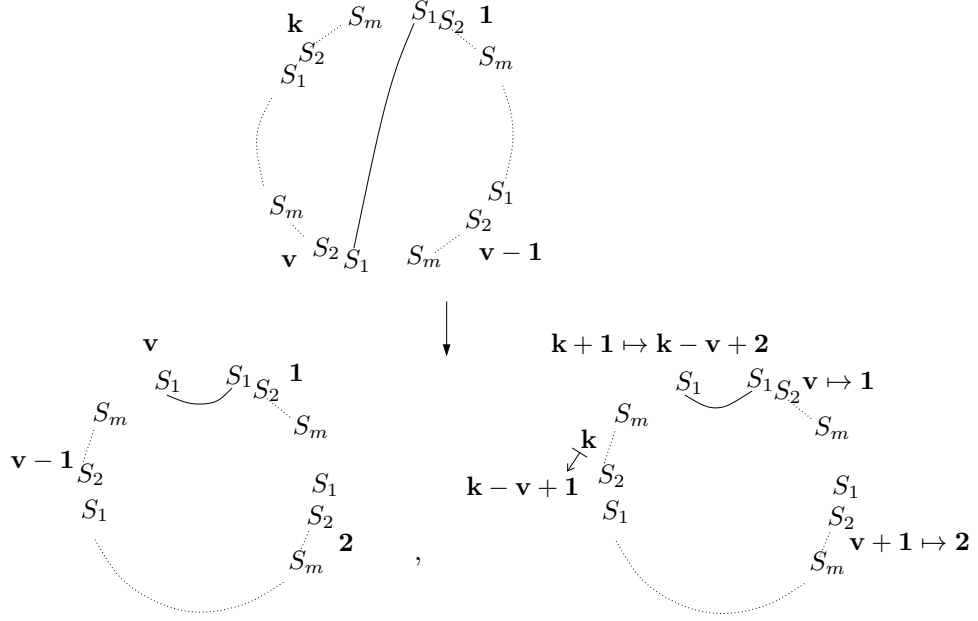
Lemma 4.6. [15, Eq. 5.63]

Let $n, r, s, t \in \mathbb{Z}$, with $n \geq 0$. Then we have:

$$\sum_{k=0}^n \frac{r}{tk+r} \cdot \frac{s}{t(n-k)+s} \binom{tk+r}{k} \binom{t(n-k)+s}{n-k} = \frac{r+s}{tn+r+s} \binom{tn+r+s}{n}$$

(If a denominator factor vanishes, the formula still makes sense by cancelling it with an appropriate factor in the numerator of a binomial coefficient.) \square

Note that the formula holds in greater generality but we shall not need it. We now have an alternative proof of Corollary 3.8:

FIGURE 10. The case with an arc incident with S_1 in vertex 1.

Theorem 4.7. Fix integers $m \geq 1, k \geq 0$. The cardinality of the set of connected noncrossing RNA m -diagrams of degree k is given by

$$T_{k,m} = \frac{m}{(m-2)k+2} \binom{(m-1)k}{k-1}.$$

Proof. By Corollary 4.5 we have that

$$T_{k,m} = \sum_{v=0}^{k-1} S_{v,m} S_{k-v,m} = \sum_{v=0}^k S_{v,m} S_{k-v,m} - S_{k,m},$$

since $S_{0,m} = 1$. Hence, using Lemma 4.6, we have:

$$\begin{aligned}
T_{k,m} &= \sum_{v=0}^k \frac{1}{(m-1)v+1} \binom{(m-1)v+1}{v} \frac{1}{(m-1)(k-v)+1} \binom{(m-1)(k-v)+1}{k-v} \\
&\quad - S_{k,m} \\
&= \frac{2}{(m-1)k+2} \binom{(m-1)k+2}{k} - \frac{1}{(m-1)k+1} \binom{(m-1)k+1}{k} \\
&= \frac{2}{(m-1)k+2} \cdot \frac{(m-1)k+2}{(m-2)k+2} \binom{(m-1)k+1}{k} - \frac{1}{(m-1)k+1} \binom{(m-1)k+1}{k} \\
&= \frac{2}{(m-2)k+2} \binom{(m-1)k+1}{k} - \frac{1}{(m-1)k+1} \binom{(m-1)k+1}{k} \\
&= \left(\frac{2}{(m-2)k+2} - \frac{1}{(m-1)k+1} \right) \binom{(m-1)k+1}{k} \\
&= \left(\frac{2}{(m-2)k+2} - \frac{1}{(m-1)k+1} \right) \frac{(m-1)k+1}{k} \binom{(m-1)k}{k-1} \\
&= \frac{((m-1)k+1)(2((m-1)k+1) - ((m-2)k+2))}{k((m-2)k+2)((m-1)k+1)} \binom{(m-1)k}{k-1} \\
&= \frac{m}{(m-2)k+2} \binom{(m-1)k}{k-1},
\end{aligned}$$

as required. \square

5. CONVOLUTION

In this section we use a similar technique to show that the sequence $T_{k,m}$, $k = 1, 2, \dots$ can be regarded as an m -fold convolution of the sequence $S_{k,m}$, $k = 0, 1, 2, \dots$. We fix a symbol S_r in the following.

Lemma 5.1. *The following sets are in bijection:*

The set of connected noncrossing RNA m -diagrams of degree $k+1$ in which there is an arc between vertex 1 and vertex $k+1$ with symbol S_r .

\updownarrow

The set of connected noncrossing RNA m -diagrams of degree $k+1$ in which there is an arc between vertex 1 and vertex $k+1$ with symbol S_1 .

Proof. Given a structure in the first set, note that there can be no arcs incident with vertex 1 with symbol S_t with $t < r$ and no arcs with vertex $k+1$ with symbol S_t with $t > r$ (since the structure is noncrossing). Similarly, given a structure in the second set, there can be no arcs incident with vertex $k+1$ with symbol S_t with $t > 1$. It follows that moving an element Σ of the first set $r-1$ steps anticlockwise, each step moving each arc one symbol anticlockwise in the diagram, gives a noncrossing RNA m -diagram Σ' of degree $k+1$ in which there is an arc between vertex 1 and vertex $k+1$ with symbol S_1 . Since Σ' has the same number of arcs, it must be connected by Remark 3.3(1). Hence Σ' is a structure in the second set and we get a map from the first set to the second set. It is clear that the inverse of this map is shifting $r-1$ steps clockwise. \square

We note that removing symbols S_2, S_3, \dots, S_m from vertex $k+1$ from an element of the second set makes no difference, since vertex $k+1$ can only have an arc to vertex 1 with symbol S_1 by the noncrossing condition.

Corollary 5.2. *The number of connected noncrossing RNA m -diagrams of degree $k+1$ on m symbols in which there is an arc between vertex 1 and vertex $k+1$ with symbol S_r is equal to $S_{k,m}$.*

Proof. Use Lemma 5.1 and Theorem 4.1, noting that an object as in Theorem 4.1(a) must have an arc between vertex $k + 1$ and vertex 1 with symbol S_1 by the non-crossing and connectedness conditions. \square

Proposition 5.3. *Let $S_{k,m}, T_{k,m}$ be as above and set $k \geq 1$. Then*

$$T_{k,m} = \sum_{k_1 \geq 0, \dots, k_m \geq 0, k_1 + \dots + k_m = k-1} S_{k_1,m} S_{k_2,m} \cdots S_{k_m,m},$$

i.e. the sequence $T_{1,m}, T_{2,m}, \dots$ is the m -fold convolution of the sequence $S_{0,m}, S_{1,m}, \dots$

Proof. Let Σ be a connected noncrossing RNA m -diagram with k vertices, with corresponding labelled m -edge-coloured tree G . Recall (see Remark 3.3) that $\sigma_G(1) = k$, i.e. applying the symbols S_1, S_2, \dots, S_m in order to the vertex 1 takes us from 1 to k . Suppose that applying symbol S_j takes us from vertex i to vertex $i + k_j$ for $j = 1, 2, \dots, m$. Note that $k_j \geq 0$ for all j by the noncrossing condition. We see that Σ corresponds to the joining of m connected noncrossing RNA m -diagrams $\Sigma_1, \dots, \Sigma_m$, where Σ_j has degree $k_j + 1$, and in which there is an arc between vertex 1 and vertex $k_j + 1$ in Σ_j with colour S_j . Note that we must have $k_1 + \dots + k_m = k - 1$. Thus we have a bijection between the set of connected noncrossing RNA m -diagrams of degree k and the set of such m -tuples. The formula in the proposition follows from this and Corollary 5.2. \square

Remark 5.4. We saw in Theorem 4.2 that the sequence $S_{0,m}, S_{1,m}, \dots$ is the Fuss-Catalan sequence of degree $m - 1$. Thus we have shown that the sequence $T_{1,m}, T_{2,m}, \dots$ is the m -fold convolution of the Fuss-Catalan sequence of degree $m - 1$. This fact appears in [20, §3]. See also [30] (sequences A071724 and A006629) for discussion of the cases $m = 3$ and $m = 4$.

6. THE TOTAL NUMBER OF TREES OF RELATIONS

In Section 3 we counted the number of labelled m -edge-coloured trees with k vertices and a fixed circular order. We now would like to count the total number of labelled m -edge-coloured trees with k vertices, i.e. with arbitrary circular order. By Remark 3.10, the number of labelled m -edge-coloured trees with k vertices and fixed circular order σ (where σ is a k -cycle) is given by

$$T_{k,m} = \frac{m}{(m-2)k+2} \binom{(m-1)k}{k-1}$$

We also have:

Lemma 6.1. *Let G be a labelled m -edge-coloured tree with k vertices. Then the circular order σ_G of G is a k -cycle.*

Proof. This is clearly true if $k = 1$ or 2 . Suppose it holds for smaller values of k . Let v be a vertex of G which is not a leaf. Suppose that v is incident with edges e_1, e_2, \dots, e_d in G , coloured with $S_{r_1}, S_{r_2}, \dots, S_{r_d}$ respectively, where $r_1 < r_2 < \dots < r_d$. Let the end-points of these edges (other than v) be v_1, v_2, \dots, v_d . Removing v from G leaves d subtrees G'_1, G'_2, \dots, G'_d incident with v_1, v_2, \dots, v_d respectively. Let G_i be the subtree G'_i with v and e_i reattached to v_i .

By the inductive hypothesis the σ_{G_i} are all cycles. Hence, repeatedly applying σ_{G_1} to v cycles through the vertices of G_1 . Since $\sigma_G = \sigma_{G_1}$ on all vertices of G_1 except $w = \sigma_{G_1}^{-1}(v)$, repeatedly applying σ_G also cycles through all the vertices of G_1 . Since r_2 is minimal such that S_{r_2} is a symbol colouring an edge incident with v with $r_2 > r_1$, $\sigma_G(w)$ will lie in G'_2 . In fact $\sigma_G(w) = \sigma_{G_2}(v)$, since in G_2 , v is not incident with any edge with symbol other than S_{r_2} . Repeatedly applying σ_G then

cycles through the vertices of G'_2 before coming back to $\sigma_{G'_2}^{-1}(v)$. Repeating this argument, we see that repeatedly applying σ_G to v first cycles through G'_1 , then through G'_2, G'_3, \dots, G'_d in order before eventually returning to v . \square

(Note that it follows that the circular order of an m -edge-coloured tree is also a cycle.) We thus see that the number of possible circular orders of labelled m -edge-coloured trees with k vertices is the number of k -cycles in \mathfrak{S}_k , i.e. $(k-1)!$. It follows that:

Proposition 6.2. *The total number of labelled m -edge-coloured trees of relations with k vertices is:*

$$U_{k,m} = \frac{m((m-1)k)!}{((m-2)k+2)!}$$

Proof. By the above discussion and Lemma 6.1, we have:

$$\begin{aligned} U_{k,m} &= T_{k,m}(k-1)! \\ &= \frac{m}{((m-2)k+2)} \frac{((m-1)k)!}{(k-1)!((m-2)k+1)!} (k-1)! \\ &= \frac{m((m-1)k)!}{((m-2)k+2)!}. \end{aligned}$$

\square

Remark 6.3. This is already known, and appears in [32, 5.28, p124], credited to I. Gessel. Stanley asked whether there was a simple bijective proof of this fact. A bijective proof was subsequently given in [6]. We have thus obtained here a new bijective proof of this fact. Note that $U_{k,m}$ can be rewritten as:

$$U_{k,m} = \frac{m(k-2)!((m-1)k)!}{((m-1)k-(k-2))!(k-2)!} = m(k-2)! \binom{(m-1)k}{k-2}.$$

Example 6.4. For $m = 3, 4, 5, 6$, we give below some values of $U_{k,m}$ for small k .

k	1	2	3	4	5	6
$U_{k,3}$	1	3	18	168	2160	35640
$U_{k,4}$	1	4	36	528	10920	293760
$U_{k,5}$	1	5	60	1200	34200	1275120
$U_{k,6}$	1	6	90	2280	82800	3946320

Putting together the results of the previous sections with the above discussion, we have:

Corollary 6.5. *There are bijections between the following sets:*

- (1) *The set of pairs consisting of a rooted diagonal-coloured m -angulation of degree k up to rotation and a k -cycle;*
- (2) *The set of $(m$ -gon)-labelled, diagonal-coloured m -angulations with k m -gons up to rotation;*
- (3) *The set of labelled m -edge-coloured trees with k vertices.*

All of the above sets have cardinality:

$$U_{k,m} = \frac{m((m-1)k)!}{((m-2)k+2)!}.$$

Proof. The set of labelled m -edge-coloured trees with k vertices and a fixed circular order is in bijection with the set of rooted diagonal-coloured m -angulations of degree k up to rotation by Corollary 3.7 (see also Remark 3.10). Thus mapping such a

tree to its corresponding m -angulation together with the circular order of the tree gives a bijection between (3) and (1), using Lemma 6.1. To go between (2) and (3), argue as in Theorem 3.6 (the m -gon-labelling of the m -angulation corresponds to the vertex-labelling of the tree). \square

7. GENERALISED INDUCTION

In this section we give the definition of our generalised induction on labelled m -edge-coloured trees with k vertices, generalising the induction in the case $m = 3$ defined in [2]. The induction in [2] leads in [9] to the construction of new languages generalising the Sturmian languages, used to study k -interval exchange transformations. Induction generates new labelled m -edge-coloured trees with the same number of vertices starting with a given such tree, and the transitive closure is an equivalence relation. We show that the circular order is an invariant, giving rise to a classification of the equivalence classes by k -cycles in \mathfrak{S}_k .

Given an labelled m -edge-coloured tree G and integers $i, j \in \{1, \dots, m\}$ we define a maximal $S_i - S_j$ chain B in G to be a (linear) subtree of G whose edges are only coloured S_i and S_j such that no other edges incident to B are coloured by S_i or S_j .

Definition 7.1. Let G be a labelled m -edge-coloured tree with k vertices. Fix $i, j \in \{1, \dots, m\}$ with $i < j$. Let B be a maximal $S_i - S_j$ chain in G . Define $R_{i,j}^B(G)$ to be the labelled m -edge-coloured tree with k vertices obtained from G by

- first removing all subtrees in the complement of the maximal chain B
- interchanging the vertices of each edge of B coloured by S_j
- interchanging the symbols S_i and S_j on the whole maximal chain B
- reattaching the previously removed subtrees to B at the vertices with the same label they were removed from.

Similarly, define $L_{i,j}^B(G)$, where in the second bullet point in the above definition we interchange the vertices of each edge coloured by S_i rather than those coloured by S_j . We also set $R_i^B := R_{i,i+1}^B$ and $L_i^B := L_{i,i+1}^B$ and we will write R_i and L_i if B is clear from the context.

Lemma 7.2. Let $i, j \in \{1, \dots, m\}$ with $i < j$ and B be a maximal $S_i - S_j$ -chain with no incident edges coloured by S_{i+1}, \dots, S_{j-1} . Then the induction $R_{i,j}^B(G)$ is a product of inductions of the form R_l for $l = i, i+1, \dots, j-1$.

Proof. Suppose first that B has the following form:

$$a_1 \xrightarrow{S_i} a_2 \xrightarrow{S_j} a_3 \xrightarrow{S_i} a_4 \quad \dots \quad a_{r-1} \xrightarrow{S_i} a_r .$$

Then $R_{i,j}^B(G)$ is the chain

$$B' = a_1 \xrightarrow{S_j} a_3 \xrightarrow{S_i} a_2 \xrightarrow{S_j} a_5 \xrightarrow{S_i} a_4 \quad \dots \quad a_{r-2} \xrightarrow{S_j} a_r .$$

On the other hand, applying $R_i, R_{i+1}, \dots, R_{j-2}$, in order (in each case to all the maximal chains of appropriate type contained in B), we obtain the maxi-

mal $S_{j-1} - S_j$ -chain $a_1 \xrightarrow{S_{j-1}} a_2 \xrightarrow{S_j} a_3 \xrightarrow{S_{j-1}} a_4 \quad \dots \quad a_{r-1} \xrightarrow{S_{j-1}} a_r$.

Next, apply R_{j-1} to the whole chain B . Then, applying $R_{j-2}, R_{j-3}, \dots, R_i$ in decreasing order (in each case to all the maximal chains of appropriate type contained in B) gives the chain B' .

The proof works in a similar way for the other configurations of maximal $S_i - S_j$ -chains, i.e. the chains of the form

$$a_1 \xrightarrow{S_i} a_2 \xrightarrow{S_j} a_3 \xrightarrow{S_i} a_4 \quad \dots \quad a_{r-1} \xrightarrow{S_j} a_r$$

and

$$a_1 \xrightarrow{S_j} a_2 \xrightarrow{S_i} a_3 \xrightarrow{S_j} a_4 \quad \dots \quad a_{r-1} \xrightarrow{S_j} a_r .$$

□

Remark 7.3. (1) In the above proof, we may replace the inductions R_l with inductions L_l for $l = i, i + 1, \dots, j - 1$.

(2) If we replace R_{j-1} with L_{j-1} in the above proof we obtain the induction $L_{i,j}^B(G)$ instead.

(3) Induction can also be defined on m -edge-coloured trees: For an m -edge-coloured tree with k vertices, choose an arbitrary vertex-labelling, apply induction, and then remove the vertex labelling. It is clear that this is independent of the vertex-labelling chosen.

(4) The inductions $R_{i,j}^B$ and $L_{i,j}^B$ are mutually inverse maps.

Definition 7.4. Call two labelled m -edge-coloured trees with k vertices G, G' *induction equivalent* if there is a sequence of inductions taking G to G' , either of the form L_i , for $2 \leq i \leq m - 1$ or of the form R_i , for $1 \leq i \leq m - 2$. Clearly this is a reflexive relation. It is symmetric since L_i and R_i are inverse maps and it is easy to see that it is transitive. Hence it is an equivalence relation. We write $\Gamma(G)$ for the equivalence class containing G .

The inductions R_i and L_i and the above notion of induction equivalence are the ones that behave well with respect to the circular order, as we shall see below. However, we shall also denote by $\Gamma_{gen}(G)$ the equivalence class of G under the equivalence given by the more general inductions $R_{i,j}$ and $L_{i,j}$.

Proposition 7.5. *Let G be a labelled m -edge-coloured tree with k vertices. Then there exists a labelled m -edge-coloured tree with k vertices whose edges are only coloured by S_1 and S_m , and a sequence of inductions, each of the form $R_{i,j}$ (with $j \leq m - 1$) or of the form $L_{i,j}$ (with $i \geq 2$) taking G to G_* .*

Proof. We first show that there exists a labelled m -edge-coloured tree G_2 with k vertices that is induction equivalent to G , none of whose edges is coloured with S_2 , by removing the symbols S_2 one by one. Firstly, remove all edges coloured with symbols S_4, \dots, S_m and call the resulting labelled 3-edge-coloured tree \tilde{G} . By [2, Prop. 5.2] there exists a sequence of inductions of the form $R_{1,2}$ and $L_{2,3}$ taking \tilde{G} to a labelled 3-edge-coloured tree \tilde{G}_2 with no edge coloured S_2 . Let G_2 be the labelled m -edge-coloured tree \tilde{G}_2 with the detached edges reattached (to the vertices with the same label). Since none of the detached edges are coloured with S_1, S_2 , or S_3 this sequence of inductions also takes G to G_2 by identifying maximal chains in \tilde{G} with corresponding maximal chains in G .

Suppose we have shown that G is induction equivalent to G_{l-1} , where G_{l-1} has no edges coloured S_2, \dots, S_{l-1} .

Then start by detaching all edges coloured S_{l+2}, \dots, S_m . Call the resulting labelled m -edge-coloured tree \tilde{G} . By [2, Prop. 5.2] there is a sequence of inductions of the form $R_{1,l}$ and $L_{l,l+1}$ taking \tilde{G} to \tilde{G}_l , where \tilde{G}_l has no edges coloured S_2, \dots, S_l . Reattach the detached edges and call the resulting labelled m -edge-coloured tree G_l . Since none of the reattached edges are coloured by S_1, S_l , or S_{l+1} , this sequence of inductions takes G_{l-1} to G_l by identifying the maximal chains in \tilde{G} with the maximal chains in G_{l-1} . Note that none of the symbols S_2, \dots, S_l appears in G_l . Hence, by induction on l , we can construct $G_{m-1} = G_*$ with no symbols S_2, \dots, S_{m-1} and a sequence of inductions, each of form R_{ij} (with $j \leq m - 1$) or L_{ij} (with $i \geq 2$) taking G to G_* . □

Remark 7.6. In the above proof, the inductions $R_{1,l}$ are applied in a situation satisfying the hypotheses of Lemma 7.2, and hence each can be written as a product of the inductions R_1, \dots, R_{l-1} .

Corollary 7.7. *Let G be a labelled m -edge-coloured tree with k vertices. Then there exists a labelled m -edge-coloured tree with k vertices, G_* , containing only symbols S_1 and S_m and a sequence of inductions, each of the form R_i or L_i , taking G to G_* .*

Corollary 7.8. *Let G be a labelled m -edge-coloured tree with k vertices. Then every possible m -edge-coloured tree with k vertices appears as the underlying unlabelled m -edge-coloured tree of a labelled m -edge-coloured tree in $\Gamma(G)$.*

Proof. Let \mathcal{S} and \mathcal{S}' be arbitrary m -edge-coloured trees with k vertices. Then, by Corollary 7.7, \mathcal{S} is induction equivalent to an m -edge-coloured tree with k vertices containing only the symbols S_1 and S_m ; similarly for \mathcal{S}' (see Remark 7.3(3)). If k is odd there is only one such tree:

$$\bullet \xrightarrow{S_m} \bullet \xrightarrow{S_1} \bullet \xrightarrow{S_m} \bullet \quad \dots \quad \bullet \xrightarrow{S_1} \bullet,$$

and it follows that \mathcal{S} and \mathcal{S}' are inductively equivalent. If k is even, there are two such trees:

$$\begin{aligned} \mathcal{S}_m : & \bullet \xrightarrow{S_m} \bullet \xrightarrow{S_1} \bullet \xrightarrow{S_m} \bullet \quad \dots \quad \bullet \xrightarrow{S_m} \bullet \\ \mathcal{S}_1 : & \bullet \xrightarrow{S_1} \bullet \xrightarrow{S_m} \bullet \xrightarrow{S_1} \bullet \quad \dots \quad \bullet \xrightarrow{S_1} \bullet. \end{aligned}$$

Then $R_{1,m}^{S_m}(\mathcal{S}_m) = \mathcal{S}_1$, so \mathcal{S}_m is induction equivalent to \mathcal{S}_1 by Lemma 7.2. It follows that \mathcal{S} and \mathcal{S}' are induction equivalent in this case also.

Given a labelled m -edge-coloured tree with k vertices, G , whose underlying unlabelled tree is \mathcal{S} , and an arbitrary m -edge-coloured tree with k vertices, \mathcal{S}' , the above shows that \mathcal{S} and \mathcal{S}' are inductively equivalent. It follows that G and a vertex-labelling of \mathcal{S}' are inductively equivalent, and the result follows. \square

Remark 7.9. It follows from the above proof that every possible m -edge-coloured tree with k vertices appears as the underlying unlabelled tree of a tree in $\Gamma_{Gen}(G)$.

Corollary 7.10. *If there is a sequence of general inductions of the form $R_{i,j}$ and $L_{i,j}$ between two m -edge-coloured trees with k vertices then they are induction equivalent (i.e. equivalent under L_i and R_i induction).*

Proof. This follows immediately from Corollary 7.8, since in fact any two m -edge-coloured trees with k vertices are induction equivalent. \square

We shall see below that the corresponding result does not hold for labelled trees (see Remark 7.15).

Lemma 7.11. (a) *Let k be odd, let $i \neq j$ and let G be a labelled m -edge-coloured tree with k vertices and whose edges are coloured with S_i and S_j only. Let \mathcal{S} be the underlying unlabelled tree of G . Applying $R_{i,j}$ induction on G has order k , producing k distinct trees with shape \mathcal{S} .*

(b) *Let k be even, $i \neq j$ and let G be a labelled m -edge-coloured tree with k vertices and whose edges are coloured with S_i and S_j only. Let \mathcal{S} be the underlying unlabelled tree of G . Applying $R_{i,j}$ induction to G has order k , producing $k/2$ distinct labelled trees whose underlying unlabelled tree is \mathcal{S} in each case and $k/2$ distinct labelled trees whose underlying unlabelled tree is $R_{i,j}(\mathcal{S})$.*

Proof. (a) Since \mathcal{S} contains only the symbols S_i and S_j , it is a line. Suppose the line is drawn horizontally and suppose the leftmost edge has colour S_j and vertices a_1 and a_2 from left to right. Then in $R_{i,j}^d(G)$, if it is drawn with orientation given by $\bullet \xrightarrow{S_i} a_1 \xrightarrow{S_j} \bullet$ (where one of the edges may not exist), the vertex a_1 is the

d^{th} vertex from the left. It is clear that all the induced trees $R_{i,j}^d(G)$ have the same underlying unlabelled tree (those for d odd should be read from right to left).

(b) The proof is similar to the one in (a), except that for d odd, the underlying unlabelled tree of $(R_{i,j}(G))^d$ is $R_{i,j}(\mathcal{S})$. \square

Remark 7.12. (1) Since in the context of Lemma 7.11 the underlying unlabelled tree of G is a line, we can replace $R_{i,j}$ in Lemma 7.11 by $R_i \dots R_{j-2} R_{j-1} R_{j-2} \dots R_i$.

(2) If we replace the R induction in Lemma 7.11 by L induction, the result holds and is proved by a similar argument.

Lemma 7.13. *Let G be a labelled m -edge-coloured tree with k vertices containing a maximal S_i - S_j chain B with no incident edges coloured S_l , $i < l < j$. Then the circular order, σ_G , is unchanged after $R_{i,j}$ or $L_{i,j}$ induction on B is applied.*

Proof. We show that the circular order is invariant under $R_{i,j}$ -induction in the context given (the proof for $L_{i,j}$ -induction follows a similar argument).

Let G be a labelled m -edge-coloured tree with k vertices and circular order σ_G and containing a maximal S_i - S_j -chain B such that no edges coloured S_l , for $i < l < j$ are incident to B . Let $G' = R_{i,j}^B(G)$ with circular order $\sigma_{G'}$. Let $i < j$ and let a be a vertex in B . Let S'_1, \dots, S'_m denote the maps corresponding to the symbols S_1, \dots, S_m in the labelled m -edge-coloured tree G' . Then $S_j S_i S_j(a) = S'_j(a)$, $S_j(a) = S'_j(a)$ and, for $l \neq j$, $S_l(a) = S'_l(a)$. Similarly, after applying $L_{i,j}$ -induction, we have $S_i S_j S_i(a) = S'_i(a)$, $S_i(a) = S'_i(a)$ and, for $l \neq i$, $S_l(a) = S'_l(a)$. There are three possible situations to consider.

Case 1: Suppose first that a has an incident edge coloured with S_l with $1 \leq l < i$. We assume that l is minimal. Let T be the subtree of $G \setminus B$ connected to a via this edge. Then $\sigma_G(a)$ lies in T . Applying the induction $R_{i,j}^B$ to G results in an m -edge-coloured tree G' in which a is reconnected to T by the edge coloured with S_k . Therefore $\sigma_G(a) = \sigma_{G'}(a)$.

Case 2: Suppose that a is not incident with any edges coloured with S_r for $r < i$, but that the vertex $S_j S_i(a)$ has an incident edge coloured with the S_l where $j < l \leq m$. We take l minimal. Let T be the subtree of $G \setminus B$ connected to $S_j S_i(a)$ via this edge. Then $\sigma_G(a)$ lies on the subtree T . Applying $R_{i,j}^B(G)$ results in an m -edge-coloured tree G' such T is reconnected, via the edge coloured S_l , to $S'_j S'_i(a) = S_j S_i(a)$. Therefore $\sigma_G(a) = \sigma_{G'}(a)$.

Case 3: Suppose that neither a nor $S_j S_i(a)$ has any incident edges coloured S_l , for $l \neq i, j$. Then $S_l(a) = a$ for all $l \neq i, j$ and $S_l(S_j S_i(a)) = S_j S_i(a)$ for all $l \neq i, j$. By our assumptions, for $i < l < j$, we have that $S'_l = S_l$ on the whole maximal chain. Hence, using the relations from the beginning of the proof,

$$\begin{aligned} \sigma_{G'}(a) &= S'_m \cdots S'_1(a) \\ &= S'_m \cdots S'_{j+1} S'_j S'_i(a) \\ &= S'_m \cdots S'_{j+1} S_j S_i(a) \\ &= S_j S_i(a) \\ &= S_m \cdots S_{j+1} S_j S_i(a) \\ &= S_m \cdots S_1(a). \end{aligned}$$

\square

Corollary 7.14. *The circular order of a labelled m -edge-coloured tree with k vertices is invariant under R_i and L_i induction.*

Proof. In Lemma 7.13 assume that $j = i + 1$. \square

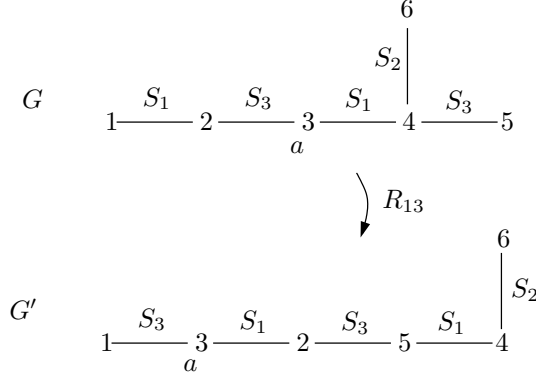


FIGURE 11. R_{13} -induction does not preserve the circular order:
 $\sigma_G(3) = 6$ while $\sigma_{G'}(3) = 5$

Remark 7.15. In general, the circular order of a labelled m -edge-coloured tree with k vertices is not invariant under general $R_{i,j}$ and $L_{i,j}$ induction for $i < j + 1$: in the above proof suppose that we have an edge e coloured S_l , for $i < l < j$ connected to $S_i(a)$ in the chain B of G and let T be a subtree of G connected to e . Then $\sigma_G(a)$ lies on the subtree T . On the other hand in $R_{i,j}^B(G)$, the edge S_k is connected to $S_i(a) = S'_i S'_j S'_i(a)$ and thus $\sigma_{G'}(a) \neq \sigma_G(a)$ in general. For an example of this with $i = 1$ and $j = 3$ (and T just consisting of one vertex), see Figure 11.

Note that this means that it is not possible, in general, to write R_{ij} as a composition of inductions of form R_i and L_i , despite Lemma 7.2 (which says that sometimes this is possible).

We finally obtain:

Theorem 7.16. *Let G and G' be labelled m -edge-coloured trees with k vertices. Then G' is in $\Gamma(G)$ if and only if $\sigma_G = \sigma_{G'}$.*

Proof. Suppose that G' is in $\Gamma(G)$. It then follows directly from Lemma 7.14 that $\sigma_G = \sigma_{G'}$. Conversely, suppose that $\sigma_G = \sigma_{G'}$. By Corollary 7.7 and the proof of Corollary 7.8 there exist labelled m -edge-coloured trees with k vertices, G_* in $\Gamma(G)$ and G'_* in $\Gamma(G')$ where the underlying unlabelled trees of G_* and G'_* contain symbols S_1 and S_m only.

Suppose that k is odd and that G_* and G'_* are as follows:

$$\begin{aligned} G_* : a_1 \xrightarrow{S_i} a_2 \xrightarrow{S_j} a_3 \xrightarrow{S_i} a_4 \quad \dots \quad a_{r-1} \xrightarrow{S_j} a_k \\ G'_* : a'_1 \xrightarrow{S_i} a'_2 \xrightarrow{S_j} a'_3 \xrightarrow{S_i} a'_4 \quad \dots \quad a'_{r-1} \xrightarrow{S_j} a'_k. \end{aligned}$$

Then $\sigma_{G_*} = (a_1 a_2 a_4 \dots a_{k-1} a_k a_{k-2} \dots a_3) = (a'_1 a'_2 a'_4 \dots a'_{k-1} a'_k a'_{k-2} \dots a'_3) = \sigma_{G'_*}$. Because of its underlying unlabelled tree, G'_* is determined by a'_1 and its circular order. Thus, given G_* there are at most k possibilities for G'_* such that $\sigma_{G_*} = \sigma_{G'_*}$ holds.

By Lemma 7.11, Remark 7.12, and Lemma 7.14, repeatedly applying either $R_{i,j}$ -induction or $L_{i,j}$ -induction to G_* gives k distinct labelled m -edge-coloured trees H with k vertices satisfying $\sigma_{G_*} = \sigma_H$. Hence any G'_* such that $\sigma_{G_*} = \sigma_{G'_*}$ must be one of these trees H . Therefore G'_* is in $\Gamma(G_*)$ and thus G' is in $\Gamma(G)$.

Suppose that k is even. In this case, there are at most $k/2$ distinct possibilities for G'_* because of the symmetry of the underlying unlabelled tree. The result then follows from Lemma 7.11, Remark 7.12, and Lemma 7.14 as above. \square

A special case of this result was proved in [2, Prop. 6.2].

8. INDUCTION ON m -ANGULATIONS OF POLYGONS

As we have seen in Section 3, the set of labelled m -edge-coloured trees with k vertices is in bijection with the set of m -gon-labelled diagonal-coloured m -angulations of a polygon P_n with $n = (m - 2)k + 2$ sides up to rotation (recall that m -gon-labelled means that the m -gons in the m -angulation are labelled with the numbers $1, 2, \dots, k$). The colours on the diagonals bounding each m -gon in the m -angulations are the symbols S_1, \dots, S_m in a clockwise order around the boundary of the m -gon. Similarly, the set of m -edge-coloured trees with k vertices is in bijection with the set of diagonal-coloured m -angulations of P_n with k m -gons up to rotation.

Our aim in this section is to rewrite induction in the language of $(m - 2)$ -clusters. We saw in Theorem 4.1 that the set of m -angulations of P_n containing k m -gons is in bijection with the set of $(m - 2)$ -clusters in the sense of Fomin-Reading [10] of type A_{k-1} . For such clusters, mutation corresponds in the m -angulation to rotating a diagonal one step anticlockwise in the subpolygon with $2m - 2$ sides obtained when the diagonal is removed. Such mutation is also well-defined on an m -angulation of P_n up to rotation. We shall see that induction has a description as a composition of such mutations (which we shall refer to as anticlockwise diagonal rotations): see Propositions 8.5 and 8.6 below.

Lemma 8.1. *Every m -angulation \mathcal{M} of a polygon has at least two m -gons with $m - 1$ boundary edges or is an m -angulation of an m -gon.*

Proof. The result is clearly true if \mathcal{M} is an m -angulation of an m -gon. Let \mathcal{M} be an m -angulation of P_n , and assume that all m -angulations of polygons with fewer sides have two boundary m -gons or are just one m -gon. Cutting \mathcal{M} along one of its m -diagonals D gives two m -angulations of polygons with fewer sides. By the induction hypothesis, each of these polygons contains an m -gon incident with its boundary and the result follows. \square

In the following two lemmas we describe how to use anticlockwise rotations of diagonals to rotate an m -angulation of P_n (containing k m -gons) anticlockwise through $2\pi/n$. Note that any clockwise rotation can easily be achieved via a composition of anticlockwise rotations; we shall therefore sometimes use clockwise rotations. We label the vertices of P_n $1, 2, \dots, n$ clockwise around the boundary, and use the notation $[i, j]$ to describe a diagonal in the polygon connecting vertex i with vertex j .

Lemma 8.2. *Let \mathcal{M} be an m -angulation of P_n by k m -gons with diagonals $[1, m]$, $[1, m + (m - 2)]$, $[1, m + 2(m - 2)]$, \dots , $[1, m + (k - 2)(m - 2)]$. Then there is an explicit sequence of diagonal rotations taking \mathcal{M} to its rotation through $2\pi/n$ anticlockwise.*

Proof. We apply the following anticlockwise diagonal rotations:

$$\begin{array}{ll}
 [1, m + (k - 2)(m - 2)] & \rightarrow [n, m + (k - 2)(m - 2) - 1] \\
 [1, m + (k - 3)(m - 2)] & \rightarrow [n, m + (k - 3)(m - 2) - 1] \\
 \vdots & \vdots \\
 [1, m + r(m - 2)] & \rightarrow [n, m + r(m - 2) - 1] \\
 \vdots & \vdots \\
 [1, m] & \rightarrow [n, m - 1]
 \end{array}$$

This produces an m -angulation \mathcal{M}' of P_n with diagonals $[n, m + (k - 2)(m - 2) - 1], [n, m + (k - 3)(m - 2) - 1], \dots, [n, m + r(m - 2) - 1], \dots, [n, m - 1]$ which corresponds to an anticlockwise rotation of \mathcal{M} through $2\pi/n$. \square

Lemma 8.3. *Let \mathcal{M} be an m -angulation of P_n containing k m -gons. Then there is an explicit sequence of diagonal rotations taking \mathcal{M} to its rotation through $2\pi/n$ anticlockwise.*

Proof. Suppose P_n has an m -angulation \mathcal{M} by k m -gons. Let M be an m -gon with $m - 1$ boundary edges and one internal edge e joining vertices $[i, i + (m - 1)]$. Let R be the union of the m -gons incident with i . Apply Lemma 8.2 to the induced m -angulation of R to rotate it one step anticlockwise. Let R' be the subpolygon R with the m -gon M' with vertices $i - 1, i, i + 1, \dots, i + (m - 1) - 1$ removed. Apply the reverse sequence to the one described in Lemma 8.2 to rotate the m -angulation of R' one step clockwise (recalling that a clockwise rotation of a diagonal coincides with a composition of anticlockwise rotations of the same diagonal). Consider the polygon Q which is given by removing the m -gon M' from P_n and transform it (using an orientation-preserving homeomorphism) to a regular $n - (m - 2)$ -gon $P_{n-(m-2)}$. Since $P_{n-(m-2)}$ has fewer sides than P_n we have inductively constructed a sequence of anticlockwise diagonal rotations rotating the m -angulation of $P_{n-(m-1)}$ anticlockwise through $2\pi/(n - (m - 1))$. Applying the corresponding sequence to the m -angulation of Q takes \mathcal{M} to its rotation anticlockwise through $2\pi/n$ as required. \square

Definition 8.4. Let \mathcal{M} be an m -angulation of P_n , diagonal-coloured with symbols S_1, S_2, \dots, S_m . If there is at least one internal diagonal in \mathcal{M} and all of the internal diagonals of \mathcal{M} are coloured only with S_i or S_j for fixed i, j , we call \mathcal{M} a *snake m -angulation*. A subpolygon of P_n with this property is called a *snake subpolygon*. Note that in any snake subpolygon the internal diagonals must be of the form $[i_1, i_2]$, $[i_2, i_3]$, and so on, and the internal angle between diagonals $[i_{r-1}, i_r]$ and $[i_r, i_{r+1}]$ must alternate between being positive and negative as r increases.

We note that such snake m -angulations appeared in the context of cluster algebras in Fomin-Zelevinsky's article [14, Sect. 3.5] (for the case $m = 3$, i.e. triangulations) and appeared for general m in [10, Sect. 5.1] under the name *m -snake*. We now show how they can be used to describe R_i induction as a sequence of $(m - 2)$ -cluster mutations, i.e. anticlockwise diagonal rotations.

Firstly, we consider induction on labelled m -edge-coloured trees with k vertices. The set of such trees is in bijection with the set of diagonal-coloured m -angulations of P_n up to rotation.

Induction on diagonal-coloured m -angulations.

Let \mathcal{M} be a diagonal-coloured m -angulation of P_n and let $1 \leq i \leq m$. Choose a maximal snake subpolygon \mathcal{B} of \mathcal{M} with internal diagonals coloured S_i or S_{i+1} . Let $\mathcal{R}_i^{\mathcal{B}}(\mathcal{M})$ be defined using the following procedure.

Step 1: It is easy to see that the induced m -angulation of \mathcal{B} contains exactly two m -gons with $m - 1$ boundary edges in \mathcal{B} .

We fix M_1 to be one such m -gon, and take e to be the internal edge of M_1 in \mathcal{B} . Let M_2 be the unique m -gon adjacent to M_1 , M_3 the unique m -gon adjacent to M_2 , and so on, with M_l the unique m -gon in \mathcal{B} adjacent to M_{l-1} for each l .

If e has colour S_i , rotate the diagonal between M_2 and M_3 one step anticlockwise, then rotate the diagonal between M_4 and M_5 one step anticlockwise and continue like this until there are 0 or 1 m -gons left in \mathcal{B} .

If e is coloured S_{i+1} , rotate the diagonals between M_1 and M_2 , M_3 and M_4 , etc. one step anticlockwise until there are 0 or 1 m -gons left in \mathcal{B} .

Exchange the colours of the edges coloured with S_i and S_{i+1} in \mathcal{B} and recolour the boundary edges in \mathcal{B} as required, using the rule that each m -gon in \mathcal{B} must have the symbols S_1, S_2, \dots, S_m clockwise on its boundary. Note that this only changes colours on the boundary edges of m -gons M in \mathcal{B} with a single internal edge in \mathcal{B} which must be coloured S_i . Connected components of the complement of \mathcal{B} in P_n incident with such m -gons M are then (temporarily) recoloured so as to be compatible with these colours.

Step 2: For each connected component C of the complement of \mathcal{B} in P_n incident with an m -gon M inside \mathcal{B} with $m - 1$ boundary edges in \mathcal{B} with internal edge e in \mathcal{B} coloured with S_{i+1} (i.e. recoloured from S_i in the above step), let $D = C \cup M$. Apply Lemma 8.3 to D to get a sequence of anticlockwise rotations rotating the induced m -angulation of D anticlockwise. Colour the diagonals of D using the colour S_i on the edge e and the clockwise rule for labelling m -gons. Note that this step has the same effect as detaching C from an edge coloured S_j , $j \neq i, i + 1$, of M and reattaching it after Step 1 to a boundary edge of M with the same symbol S_j (one step anticlockwise around the boundary of M), keeping the original labelling of C (i.e. from before Step 1). Such a boundary edge of M always exists, since $j \neq i, i + 1$. Should no such component C exist, no action is necessary.

It is easy to see that the above procedure does not depend on the initial choice of M as m -gon of \mathcal{B} with $m - 1$ boundary edges. Note also that the procedure commutes with any rotational symmetry of P_n and so gives a well defined induction on a diagonal-coloured m -angulation of P_n up to rotation which we denote the same way.

Comparing the definitions of $R_i^{\mathcal{B}}$ and $\mathcal{R}_i^{\mathcal{B}}$ it is easy to check that:

Proposition 8.5. *Suppose that $1 \leq i \leq m$ and let G be an m -edge-coloured tree with k vertices containing a maximal $S_i - S_{i+1}$ chain B . Let \mathcal{M} be the corresponding diagonal-coloured m -angulation of P_n up to rotation with maximal snake subpolygon \mathcal{B} corresponding to B . Then $R_i^{\mathcal{B}}(G)$ corresponds to $R_i^{\mathcal{B}}(\mathcal{M})$. \square*

For a simple example of Proposition 8.5, see Figure 12.

Finally, we show how the above can be modified to the case of a labelled m -edge-coloured tree with k vertices. We thus define induction on m -gon-labelled, diagonal-coloured m -angulations of P_n up to rotation. We also reduce the number of anticlockwise diagonal rotations required by allowing detaching and reattaching of subpolygons (see Step 2 above).

Induction on m -gon-labelled diagonal-coloured m -angulations.

Let \mathcal{M} be an m -gon-labelled, diagonal-coloured m -angulation of P_n and let $1 \leq i \leq m$. Choose a maximal snake subpolygon \mathcal{B} of \mathcal{M} with internal diagonals coloured S_i or S_{i+1} . Let $\mathcal{R}_i^{\mathcal{B}}(\mathcal{M})$ be defined using the following procedure.

Detach all subpolygons in the complement of \mathcal{B} in P_n . We know that the induced m -angulation of \mathcal{B} has exactly two m -gons with $m - 1$ boundary edges in \mathcal{B} . Let M_1 be one of them with internal edge e . Let M_2 be the unique m -gon adjacent to M_1 , M_3 the unique m -gon adjacent to M_2 , and so on.

If e is coloured S_i , rotate the diagonal between M_2 and M_3 one step anticlockwise. Let M'_2 and M'_3 be the new m -gons created, with M'_3 adjacent to M_1 . Repeat for M_4 and M_5 , M_6 and M_7 , etc until 0 or 1 m -gons are left.

If e is coloured S_{i+1} , rotate the diagonal between M_1 and M_2 one step anticlockwise; call the new m -gons created M'_1 and M'_2 with M'_2 incident with a boundary edge of M_1 . Repeat for M_3 and M_4 , etc. until 0 or 1 m -gons are left.

Recolour all edges in \mathcal{B} , starting by exchanging colours S_i and S_{i+1} on the diagonals of \mathcal{B} and then using the rule that every m -gon inside \mathcal{B} must have the symbols S_1, \dots, S_m clockwise on its boundary. Reattach each detached subpolygon (keeping its original diagonal labelling) to the m -gon with the same label it was

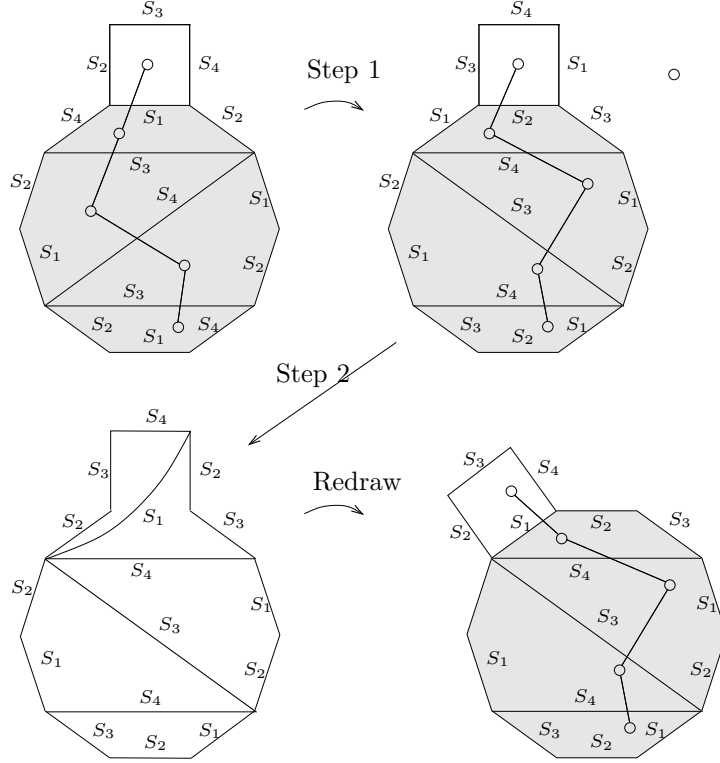


FIGURE 12. Mutations giving rise to R_3 induction (see Proposition 8.5). The snake subpolygon is shaded.

originally attached to via the edge coloured with the same symbol. The new m -gon label of M'_i is defined to be the old m -gon label of M_i ; all other m -gons retain the same m -gon label.

It is easy to see that, as in the previous case, the above procedure does not depend on the initial choice of M as m -gon of \mathcal{B} with $m - 1$ boundary edges. Also, as before, the above procedure commutes with any rotational symmetry of P_n and so gives a well defined induction on an m -gon-labelled diagonal-coloured m -angulation of P_n up to rotation which we denote the same way. As in the previous case, a comparison of the definitions of R_i^B and \mathcal{R}_i^B yields:

Proposition 8.6. *Suppose that $1 \leq i \leq m$ and let G be a labelled m -edge-coloured tree with k vertices containing a maximal $S_i - S_{i+1}$ chain B . Let \mathcal{M} be the corresponding m -gon-labelled diagonal-coloured m -angulation of P_n up to rotation with maximal snake subpolygon \mathcal{B} corresponding to B . Then $R_i^B(G)$ corresponds to $\mathcal{R}_i^B(\mathcal{M})$. \square*

Remark 8.7. For L_i induction (in both cases), it is just necessary to replace all anticlockwise rotations by clockwise rotations and vice versa.

Acknowledgement We would like to thank Luca Q. Zamboni for some helpful conversations. RJM would like to thank Sibylle Schroll and the Department of Mathematics at the University of Leicester for their kind hospitality.

REFERENCES

- [1] D. Bisch, V. Jones, *Algebras associated to intermediate subfactors*, Invent. Math. 128 (1997), no. 1, 89–157.

- [2] J. Cassaigne, S. Ferenczi and L. Q. Zamboni, *Combinatorial trees arising in the study of interval exchange transformations*, European Journal of Combinatorics 32 (2011) 14281444.
- [3] F. Chapoton, S. Fomin, and A. Zelevinsky, *Polytopal realizations of generalized associahedra*, Canad. Math. Bull. 45 (2002), no. 4, 537566.
- [4] W. Y. C. Chen, J. Qin and C. M. Reidys, *Crossings and nestings in tangled diagrams*, Electron. J. Combin. 15 (2008), no. 1, Research Paper 86, 14 pp.
- [5] W. Y. C. Chen, J. Qin, C. M. Reidys and D. Zeilberger, *Efficient counting and asymptotics of k -noncrossing tangled diagrams*, Electron. J. Combin. 16 (2009), no. 1, Research Paper 37, 8 pp.
- [6] M. Cho, D. Kim, S. Seo, and H. Shin, *Colored Prfer codes for k -edge colored trees*, Electron. J. Combin. 11 (2004), no. 1, Note 10, 7 pp.
- [7] P. Clote, *Combinatorics of saturated secondary structures of RNA*, J. Comput. Biol. 13 (2006), no. 9, 16401657.
- [8] T. Došlić, S. Dragutin and D. Veljan, *Enumerative aspects of secondary structures*, Discrete Math. 285 (2004), no. 1-3, 6782.
- [9] S. Ferenczi, L. Zamboni, *Structure of k -interval exchange transformations: induction, trajectories and distance theorems*, to appear in J. Analyse Math.
- [10] S. Fomin and N. Reading, *Generalized cluster complexes and Coxeter combinatorics*, Int. Math. Res. Not. 2005, no. 44, 2709–2757.
- [11] S. Fomin and N. Reading, *Root systems and generalized associahedra*, Geometric combinatorics, 63131, IAS/Park City Math. Ser., 13, Amer. Math. Soc., Providence, RI, 2007.
- [12] S. Fomin and A. Zelevinsky, *Cluster algebras. I. Foundations*, J. Amer. Math. Soc. 15 (2002), no. 2, 497–529
- [13] S. Fomin and A. Zelevinsky, *Cluster algebras II: Finite type classification*, Invent. Math. 154 (2003), 63121.
- [14] S. Fomin and A. Zelevinsky, *Y -systems and generalized associahedra*, Ann. of Math. (2) 158 (2003), no. 3, 977–1018.
- [15] R. L. Graham, D. E. Knuth and O. Patashnik, *Concrete mathematics. A foundation for computer science*. Second edition. Addison-Wesley Publishing Company, Reading, MA, 1994.
- [16] F. Harary, E. M. Palmer, and R. C. Read, *On the cell-growth problem for arbitrary polygons*, Discrete Math. 11 (1975), 371–389
- [17] P. Hilton and J. Pederson, *Catalan numbers, their generalization, and their uses*, Mathematical Intelligencer 13 (1991), 6475.
- [18] I. L. Hofacker, P. Schuster and P. F. Stadler, *Combinatorics of RNA secondary structures*, Discrete Applied Mathematics 88 (1998) Issues 1–3, 207–237.
- [19] I. L. Hofacker, C. M. Reidys and P. F. Stadler, *Symmetric circular matchings and RNA folding*, Discrete Mathematics (2011), doi:10.1016/j.disc.2011.06.004.
- [20] V. E. Hoggatt, Jr. and M. Bicknell, *Catalan and related sequences arising from inverses of Pascal's triangle matrices*, Fib. Quart., 14 (1976), 395–405.
- [21] F. W. D. Huang and C. M. Reidys, *Statistics of canonical RNA pseudoknot structures*, J. Theoret. Biol. 253 (2008), 570–578.
- [22] B. Keller, *Cluster algebras, quiver representations and triangulated categories*, in Triangulated Categories, editors T. Holm, P. Jorgensen and R. Rouquier, London Mathematical Society Lecture Note Series, No. 375, 2010.
- [23] R. Marsh, S. Schroll and L.Q. Zamboni, *Shapes, labelled polygonal triangulations, and the Temperley-Lieb algebra*, in preparation.
- [24] A. Markov, *Sur une question de Jean Bernoulli*, Math. Ann., vol 19., 1882, 27–36.
- [25] R. Nussinov, G. Pieczenik, J. R. Griggs and D. J. Kleitman, *Algorithms for loop matchings*, SIAM J. Appl. Math. 35 (1978), 68–82.
- [26] R. C. Penner and M. S. Waterman, *Spaces of RNA secondary structures*, Adv. Math. 101 (1993), no. 1, 3149.
- [27] C. Reidys, R. R. Wang, A. Y. Y. Zhao, *Modular, k -noncrossing diagrams* E. J. Combin. 17 (2010), 1, R76.
- [28] C. Reidys, *Combinatorial computational biology of RNA*, Pseudoknots and neutral networks. Springer, New York, 2011.
- [29] W. R. Schmitt and M. S. Waterman, *Linear trees and RNA secondary structure* Disc. Appl. Math. (1994), 51, 317–323.
- [30] N. J. A. Sloane, (2010), *The On-Line Encyclopedia of Integer Sequences*, www.research.att.com/~njas/sequences/.
- [31] L. Snyder and W. Champness, *Molecular genetics of bacteria*, Second Edition, ASM Press, USA, 2003.

- [32] R. P. Stanley, *Enumerative combinatorics*. Vol. 2, with a foreword by Gian-Carlo Rota and appendix 1 by Sergey Fomin. Cambridge Studies in Advanced Mathematics, 62. Cambridge University Press, Cambridge, 1999.
- [33] P. R. Stein and M. S. Waterman, *On some new sequences generalizing the Catalan and Motzkin numbers*, Discrete Math. 26 (1978) 261-272.
- [34] E. Tzanaki, *Polygon dissections and some generalizations of cluster complexes*, J. Combin. Theory Ser. A 113 (2006), no. 6, 1189–1198.
- [35] M. S. Waterman, *Secondary structure of single-stranded nucleic acids*, Studies in foundations and combinatorics, pp. 167-212, Adv. in Math. Suppl. Stud., 1, Academic Press, New York-London, 1978.
- [36] T. F. Smith and M. S. Waterman, *RNA secondary structure: a complete mathematical analysis*, Math. Biosci. 42 (1978), 257-266.
- [37] M. S. Waterman, *Applications of Combinatorics to Molecular Biology*, Handbook of Combinatorics, Edited by R. Graham, M. Grötschel and L. Lovász, Elsevier Science, 1995.
- [38] R. Willenbring, *RNA secondary structure, permutations, and statistics*, Discrete Appl. Math. 157 (2009), no.7, 1607-1614.
- [39] M. Zuker, *RNA Folding Prediction: The Continued Need for Interaction Between Biologists and Mathematicians*, Lectures on Mathematics in the Life Sciences 17, 87-124 (1986).

SCHOOL OF MATHEMATICS, UNIVERSITY OF LEEDS, LEEDS LS2 9JT, UNITED KINGDOM
E-mail address: `marsh@maths.leeds.ac.uk`

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF LEICESTER, UNIVERSITY ROAD, LEICESTER
 LE1 7RH, UNITED KINGDOM
E-mail address: `ss489@le.ac.uk`